# Kinetic Study of Human Thymine DNA Glycosylase and Examination of Its Interaction with Apurinic Endonuclease1 in Base Excision Repair

**Mika Abu**

**A thesis submitted to University of London;**

**Doctor of Philosophy in Biochemistry**

**September 2004**

**University College London**





ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

# 1. ABSTRACT

In the human genome, cytosine is exclusively methylated at CpG sequences to give 5-methylcytosine. The C4-amino group of 5-methylcytosine is susceptible to spontaneous hydrolysis, generating thymine in a G·T mismatch. Thymine DNA glycosylase (TDG) detects the mismatch and cleaves the glycosidic bond of thymine. Then apyrimidinic/apurinic endonuclease 1 (APEX1) displaces TDG from the abasic site and cuts the phosphodiester bond 5' to the abasic site in base excision repair. The kinetic parameters, $K_d$ and $k_2$, for TDG acting on thymine and ethenocytosine substrates, were measured. The excision of both substrates was very dependent upon the base 5' to the mismatched guanine. TDG has a strong preference for both thymine and ethenocytosine in a CpG sequence. This is understandable for thymine since G·T mismatches arise in this context. However, ethenocytosine is not exclusively formed at CpG sequences. The catalytic step ($k_2$) for CpG·T was six-fold faster than CpG·εC, but was bound, as shown by $K_d$, approximately 800-fold less tightly by TDG. This large difference in $K_d$ is probably due to the unstable structure of G·εC, which allows TDG to easily flip ethenocytosine out of the DNA. Thus, the reaction of TDG with ethenocytosine in vitro is misleading. This, together with the sequence dependence of ethenocytosine excision by TDG, means that TDG cannot be the main ethenocytosine-DNA glycosylase *in vivo*. The mechanism of TDG displacement by APEX1 was examined by looking at the involvement of the DNA and by looking for protein-protein interactions between TDG and APEX1. Although the DNA is not absolutely required for the displacement of TDG by APEX1, a weak/transient protein-protein interaction was detected between TDG and APEX1 using a pull-down assay. These results suggest that APEX1 transiently interacts with TDG bound to abasic DNA to induce a conformational change in TDG that dissociates it from the abasic site.

Table of Contents

List of Tables

## List of Figures

# 2. INTRODUCTION

# 2.1. DNA structure

Deoxyribonucleic acid (DNA) is the genetic material that decides what proteins are, what cells are, what organs are, what we are, and how we are. DNA consists of four kinds of nucleotides that each contain a β-$_D$-2-deoxyribose, one or more phosphate group, and one of four bases; adenine (A), guanine (G), thymine (T), and cytosine (C) (see Figure 1). Each nucleotide is linked to the next by a phosphodiester bridge, forming a long chain.

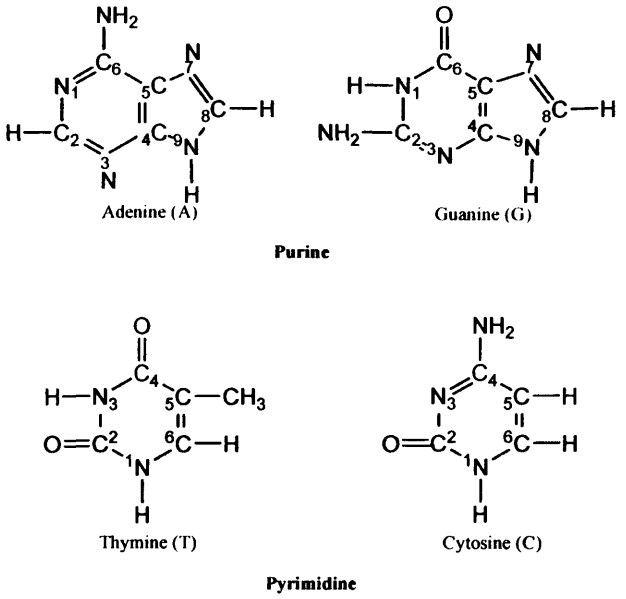



**Figure 1 Schematic structure of the four bases in DNA**

**The bases are derivatives of purine or pyrimidine. Adenine (A) and guanine (G) are purines, and thymine (T) and cytosine (C) are pyrimidines. N-9 of a purine and N-1 of a pyrimidine are attached to C-1 of deoxyribose (see Figure 2).**

In 1953, James Watson and Francis Crick (Watson and Crick, 1953) derived the structure of double helix DNA from an x-ray diffraction photograph, which depicted that two helical chains are coiled around a common axis in opposite directions. In the helix, adenine pairs with thymine and guanine pairs with cytosine, forming hydrogen bonds between them to hold the two chains together (see Figure 2 and Figure 3).

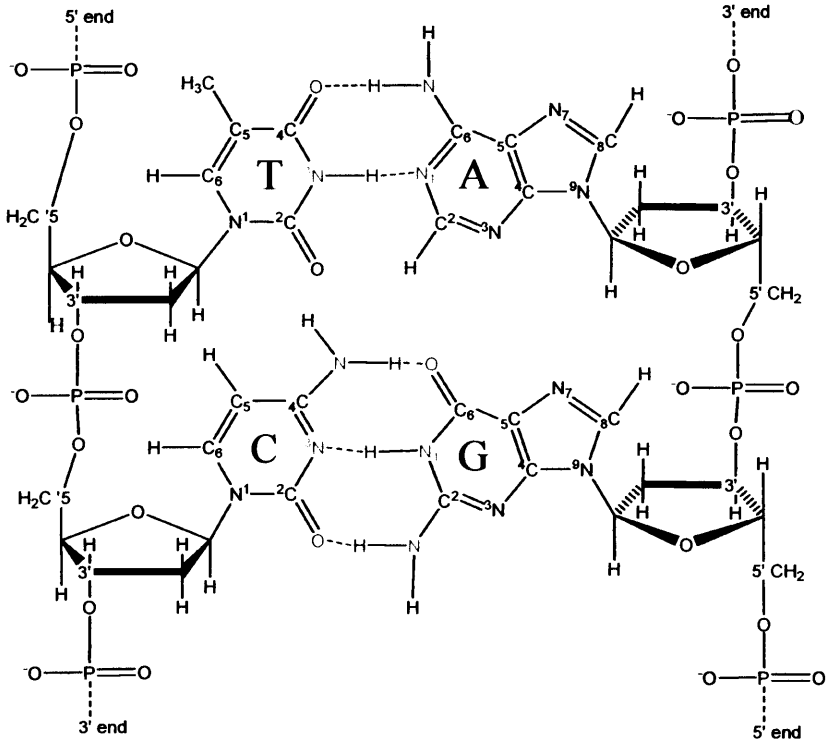



**Figure 2 Schematic structure of complementary base pairing in antiparallel double-stranded DNA**

**The phosphodiester bond is formed between the C-5' OH groups on a deoxyribose molecule and the C-3' OH group of the adjacent deoxyribose. The C-1 of the deoxyribose ring is linked to a purine or pyrimidine base. The blue dotted lines indicate the hydrogen bonds between the complementary bases.**

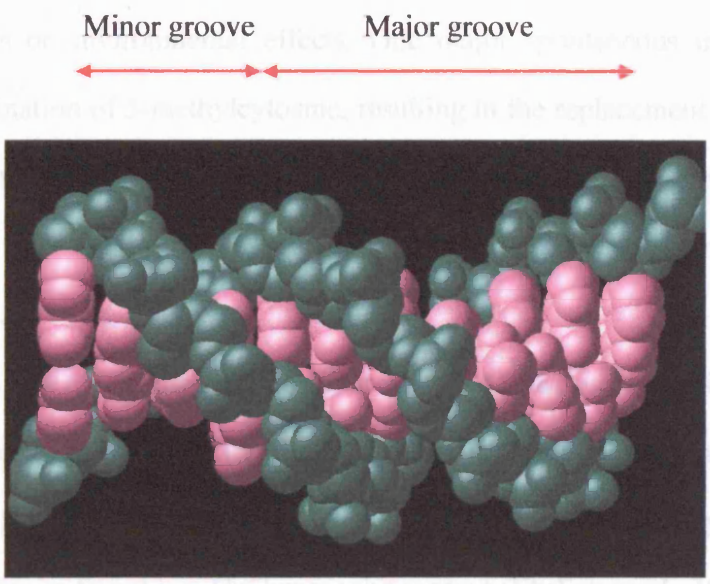



**Figure 3 Model of DNA double helix**

**The sugar-phosphate backbone is shown in green. The purine and pyrimidine bases (pink) are on the inside of the helix. The common DNA structure found in solution and in *vivo* has a major groove and a minor groove, as indicated in the figure.**

The mammalian genome contains $3.3 \times 10^9$ base pair (bp) of DNA, encoding approximately 40,000 genes. If the average coding sequence is estimated to be 1500 nucleotides (i.e. 500 amino acids), only 6 % of the genome is accounted for in coding for genes. Although, the level of gene expression is greatly varied dependent upon type of tissues, approximately 80-90 % of the mammalian genome may not be expressed at any given time. By contrast, the entire bacterial genome ($4.2 \times 10^6$ bp) is used for genes and their regulatory elements (Devlin, 1992).

The genome constantly suffers many kinds of mutations that result from normal cellular events or environmental effects. One major spontaneous mutation in DNA is deamination of 5-methylcytosine, resulting in the replacement of C by T. Thus, the sites containing 5-methylcytosine are called hotspots since they may display 10 ~ 100 times more mutations than predicted by random mutation (Lewin, 1997).

In vertebrates, ~4 % of cytosine residues in DNA bases are modified to 5-methylcytosine by DNA methyltransferases that transfer a methyl group from the methyl donor S-adenosyl-$_L$-methionine (Ehrlich *et al.*, 1982). Virtually all of the 5-methylcytosine is found in a CpG sequence context (Ehrlich *et al.*, 1982; Razin and Riggs, 1980). At 37 °C, 5-methylcytosine spontaneously deaminates to thymine at a 4-5 fold higher rate than the deamination of cytosine to uracil (Ehrlich *et al.*, 1986; Lindahl, 1993). This causes a C to T transition mutation if it is not repaired, and this transition is the most prominent type of mutation found in the p53 tumour suppressor gene (Hainaut *et al.*, 1997). However, some CpGs found in stretches of a few kilobases that account for up to 2 % of the genome (Cross and Bird, 1995), are not methylated. These unmethylated regions are called CpG islands and are found downstream of promoters in ~50 % of human genes (Ng and Bird, 1999). Methylation of DNA occurs shortly after replication during S-phase, and the pattern of methylation is tissue-specific (Razin *et al.*, 1984). The function of methylation in the genome is not clear, however, there is a correlation between methylation of CpG islands and silencing of the associated gene. Transcriptional repression is thought to happen through structural change of chromatin that may correlate to levels of histone acetylation (Bestor, 1998). In the hypothesis, methyl-CpG-binding protein 2 binds methylated DNA (Lewis *et al.*,

1992), and recruits co-repressors and a histone deacetylase to the site. Deacetylation of histones may induce tighter nucleosomal structure, and lead to reduced levels of transcription (Balylin and Herman, 2000; Bestor, 1998; Jones and Laird, 1999). Hypermethylation of CpG islands in the promoter of genes, such as hMLH1 (a DNA mismatch repair protein), BRCA1 (a gene related to familial breast and ovarian cancers), p16, and p15, are reported to be associated with inactivation of these genes in the tumour cells (Balylin and Herman, 2000).

## 2.2. DNA damage

Our genetic material is constantly challenged by endogenous and exogenous agents. Depending upon the nature of attack, a variety of modifications are induced on DNA, and different types of repair system have been developed through evolution to deal with different types of DNA damage.

### 2.2.1. Endogenous damage

The major endogenous damaging events are spontaneous hydrolysis, oxidative modification by reactive oxygen species, and alkylation of DNA.

i) Hydrolysis

The glycosidic bond of purine is liable to hydrolysis, forming an abasic site and is estimated to occur approximately 10,000 times per human cell per day (Nakamura *et al.*, 1998). The bases that contain exocyclic amino groups are also susceptible to spontaneous hydrolytic deamination, causing alteration of bases; cytosine to uracil, adenine to hypoxanthine, guanine to xanthine, and 5-methylcytosine to thymine (Figure 4).

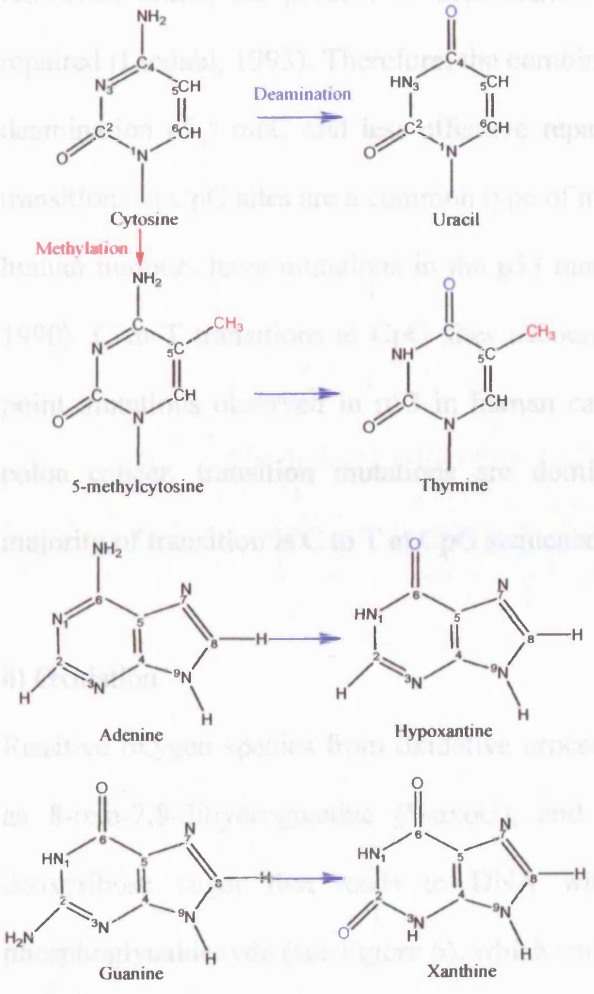



**Figure 4 Products of spontaneous deamination of the DNA bases**

**Deamination changes cytosine to uracil, 5-methylcytosine to thymine, adenine to hypoxanthine, and guanine to xanthine. Deamination of 5-methylcytosine creates thymine in a mispair with guanine.**

If the uracil formed from deamination of cytosine is not repaired, adenine is incorporated opposite it in the next round of DNA replication, leading to a C to T transition mutation. The rate of hydrolytic deamination of 5-methylcytosine is higher than that of cytosine (Ehrlich *et al.*, 1990; Lindahl, 1993; Lindahl and Nyberg, 1974) and generates thymine mismatched with guanine. Because 5-methylcytosine occurs predominantly at CpG sites, these G·T mismatches are always in a CpG context (see Figure 16). In mammalian cells, 10 % of the hydrolytic deamination may occur at 5-methylcytosine, and 90 % at cytosine. However, uracil, the product of deamination of cytosine, is far more rapidly repaired (Lindahl, 1993). Therefore, the combination of a relatively higher rate of deamination of 5-meC and less effective repair of thymine means that C to T transitions in CpG sites are a common type of mutation. Approximately half of all human tumours have mutations in the p53 tumour suppressor gene (Vogelstein, 1990). C to T transitions at CpG sites account for a quarter of all the somatic point mutations observed in p53 in human cancer (Greenblatt *et al.*, 1994). In colon cancer, transition mutations are dominant over transversion, and the majority of transition is C to T at CpG sequences in p53 (Jones *et al.*, 1992).

ii) Oxidation

Reactive oxygen species from oxidative processes generate oxidized bases such as 8-oxo-7,8-dihydroguanine (8-oxoG), and thymine glycol , and oxidized deoxyribose sugar that leads to DNA with 3'-phosphoglycolate and 3'-phosphoglycaldehyde (see Figure 6), which causes single- or double-strand DNA breaks (Pogozelski and Tullius, 1998).

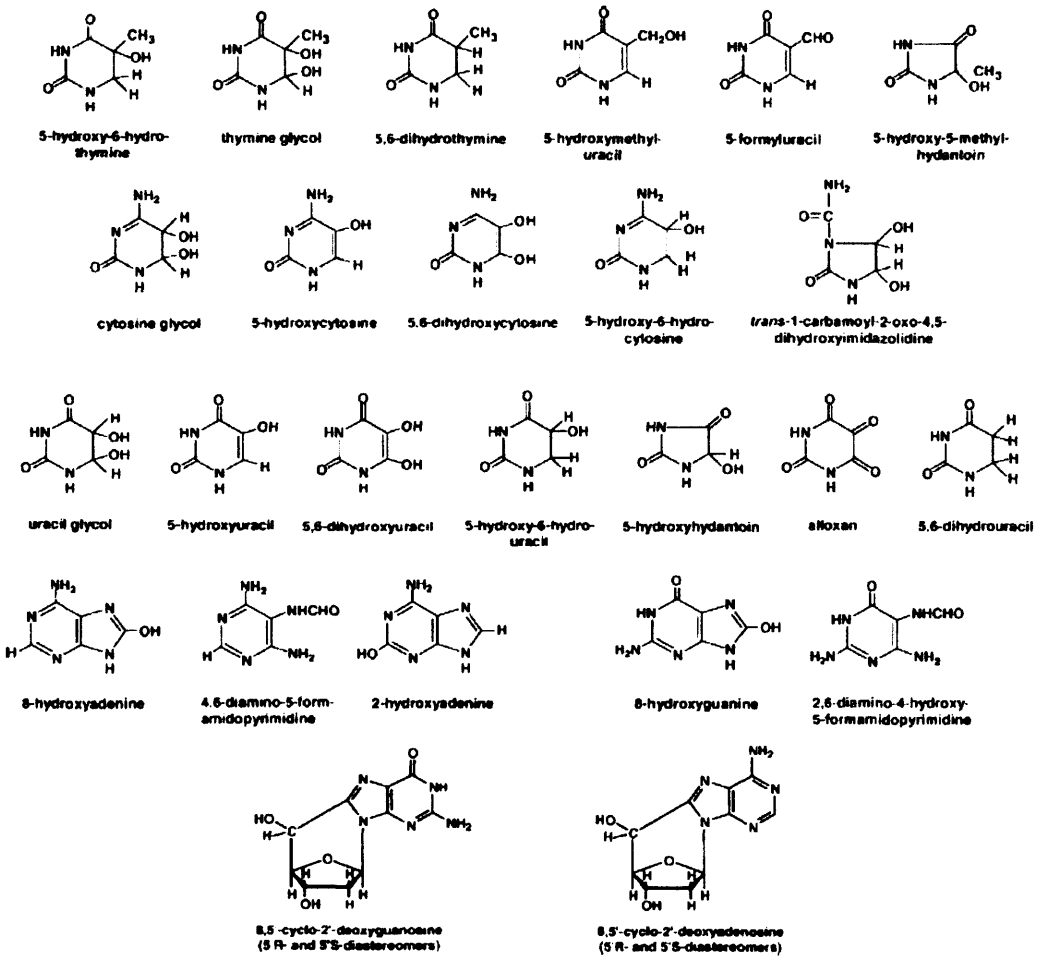



**Figure 5 Typical examples of oxidative damage of DNA**

**The figure was taken from Dizdaroglu (2003).**

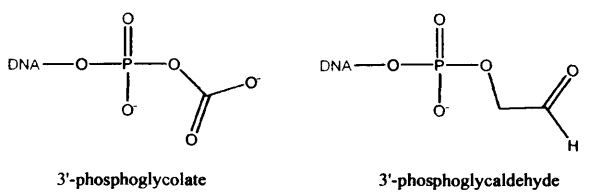



**Figure 6 Typical DNA damage caused from damage on sugar moieties by oxygen radical species (Schärer, 2003)**

If 8-oxoG is not repaired, it mispairs with adenine resulting in a G to A transition, while unrepaired thymine glycol blocks transcription (Frosina, 2000). Approximately one 8-oxo-G per $10^5$-$10^6$ nucleotides is present in mammalian cells (McCullough *et al.*, 1999).

iii) Alkylation

The cofactor of DNA methyltransferases, S-adenosylmethionine, can react with DNA to produce, for example, 3-methyladenine that is an obstacle to DNA replication (Lindahl, 2000; McCullough *et al.*, 1999; Seeberg *et al.*, 1995). Alkylation of guanine at the O-6 position to form $O^6$-methylguanine induces a mutation via incorporation of thymine opposite the $O^6$-methylguanine in successive replication (Figure 7) (Griffin *et al.*, 1994). Alkylation can also occur on the N3 or N7 position of guanine, 3-methylguanine and 7-methylguanine, respectively (Figure 7).

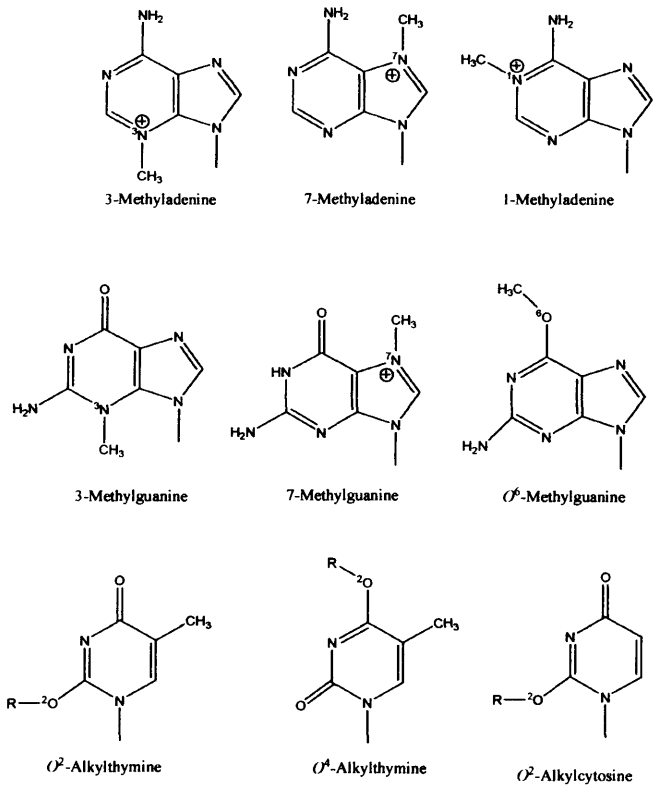



**Figure 7 Structures of typical alkylated DNA bases (Krokan *et al.*, 1997; Seeberg *et al.*, 1995)**

Products of endogenous lipid peroxidation and exposure to exogenous vinyl chloride are believed to produce the etheno (ε) adducts 3,$N^4$-ethenocytosine (εC), 1,$N^2$-ethenoguanine, $N^2$,3-ethenoguanine, and 1,$N^6$-ethenoadenine (εA) (Figure 8).

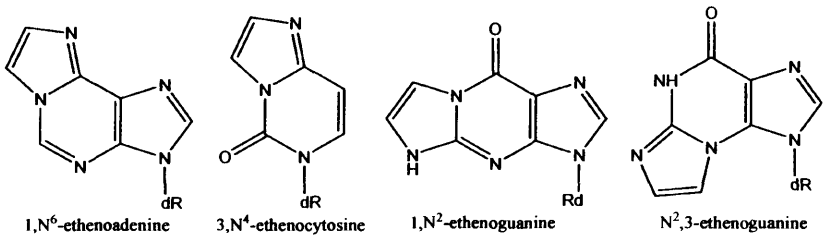



**Figure 8 Structures of etheno adducts; 1,$N^6$-ethenoadenine, 3,$N^4$-ethenocytosine, 1,$N^2$-ethenoguanine, and $N^2$,3-ethenoguanine (Singer and Hang, 1999)**

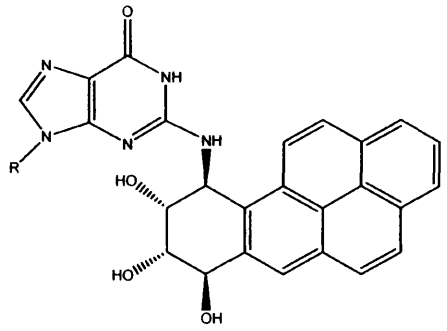
Lipid peroxidation is induced by oxidative stress caused by circumstances such as high intake of ω-6-polyunsaturated fatty acids, metal storage diseases, chronic infections and inflammations. Once oxidation of polyunsaturated fatty acids occurs, it triggers further reactions. The first oxidation creates free radicals that react with a second lipid molecule, generating a fatty acid hydroperoxide and a new free radical. One of the metabolites from the lipid peroxidation is trans-4-hydroxy-2-nonenal that is believed to be involved in the formation of εC and εA (Chung *et al.*, 1996; Ghissassi *et al.*, 1995; Nair *et al.*, 1999). Nair *et al.* (1995) estimated that εC and εA are present at 0-27 adducts per $10^9$ parent bases in human liver (Nair *et al.*, 1995). In replication, mammalian DNA polymerases preferentially incorporate A, T, and C opposite to εC (Shibutani *et al.*, 1996; Zhang *et al.*, 1995), and so εC is a mutagenic lesion.

Most of the DNA adducts formed by endogenous damaging-agents discussed above, are repaired by the base excision repair (BER) pathway (see Section 2.3.2.).

## 2.2.2. Exogenous damage

Typical exogenous agents that damage DNA include toxic compounds found in cigarette smoke, and UV radiation from sunlight. UV radiation causes formation of photoadducts between adjacent pyrimidine residues in DNA. The tobacco smoke carcinogen benzo[*a*]pyrene causes formation of benzo[a]pyrene diol epoxide-guanine (BPDE) adducts in DNA (see Figure 9). The adduct blocks DNA replication, but even if it is bypassed, it induces G to T transversions, and

these mutations are the most common mutations found in the p53 gene of lung tumours (Alekseyev and Romano, 2000). It is reported that guanine in the $^{5\text{-methyl}}$CpG sequence enhanced reactivity with BPDE (Denissenko *et al.*, 1997; Matter *et al.*, 2004). Gaseous vinyl chloride used in the plastic factories is known to cause several tumours in humans. It is activated by cytochrome P450 into chloroacetaldehyde and chloroethylene oxide that are carcinogenic metabolites. As discussed in Section 2.2.1, these compounds induce formation of eheno adducts, which leads to miscoding of DNA (Marion and Boivin-Angele, 1999) (see Figure 8).




(+)-trans-$N^2$-BPDE-dG

**Figure 9 One of Diastereomeric $N^2$-BPDE-dG Adducts (Matter *et al.*, 2004)**

Some cancer drugs, such as the cross-linking agent cisplatin, are targeted to form adducts that block replication and transcription of DNA, which then leads the cancer cells to apoptosis. In mammals, many of these modifications of DNA are repaired by nucleotide excision repair (NER; see Section 2.3.3.). Other exogenous agents, like ionising radiation, radiomimetic drugs (such as bleomycin) and reactive oxygen species induce double-strand breaks in DNA

(Friedberg *et al.*, 1995). The double-strand breaks are repaired in eukaryotes by homologous recombination or nonhomologous end joining (see Section 2.3.6.).

# 2.3. Repair of DNA damage

## 2.3.1. Direct reversal of damage

The repair of modified DNA by direct reversal is limited. In mammals, the conversion of $O^6$-methylguanine to guanine is repaired by direct reversal of the damage. $O^6$-alkylguanine alkyl transferase irreversibly transfers the methyl group from the $O^6$-methylguanine to a cysteine residue of the protein. Another example is that AlkB converts N1-methyladenine and N3-methylcytosine (Begley and Samson, 2003).

## 2.3.2. Base Excision Repair

The DNA damage generated by spontaneous hydrolysis (e.g. depurination), deamination (e.g. 5-methylcytosine to thymine mutation), and most endogenous factors are corrected by base excision repair (BER). The modified bases are initially recognized by different glycosylases dependent upon the nature of damage (see Table 1).

**Table 1 Human base excision repair proteins (Frosina, 2000; Wood *et al.*, 2001)**

| Proteins in Base Excision Repair | Function | Substrates |
|---|---|---|
| Uracil-DNA glycosylase (UDG) | Glycosylase | Uracil, 5-Hydroxyuracil, Isodialuric acid, Alloxan |
| Single strand-selective monofunctional Uracil-DNA glycosylase (SMUG1) | Glycosylase | Uracil, 5-Hydroxymethyluracil |
| Methyl-CpG-binding domain protein 4 (MBD4 / MED1) | Glycosylase, possibly endonuclease activity | Uracil, thymine, or ethenocytosine opposite guanine at CpG sequence, 5-Fluorouracil, 5-Methylcytosine in hemimethylated DNA |
| Thymine-DNA glycosylase (TDG) | Glycosylase | Uracil, thymine, or ethenocytosine opposite guanine, 5-Fluorouracil, 5-Hydroxymethyluracil |
| 8-Oxoguanine-DNA glycosylase (hOgg1) | Glycosylase and AP lyase | 8-oxoG opposite cytosine, Fragmented pyrimidines |
| Adenine-DNA glycosylase (hMyh) | Glycosylase | Adenine opposite 8-oxoG |
| A protein with a sequence similarity to E.coli Nth (endonuclease III) (hNth1) | Glycosylase and AP lyase | Thymine glycol, cytosine glycol, dihydrouracil, 5-hydroxycytosine, Fragmented pyrimidines |
| 3-methyladenine DNA glycosylase / Alkylpurine-DNA glycosylase (MPG) | Glycosylase | 3-meA, Ethenoadenine, Hypoxanthine |
| Apurinic/apyrimidinic endonuclease 1 (APEX1 / HAP1 / Ref-1) | AP endonuclease | Cleavage of phosphodiester bond at 5'-side of abasic site |
| Apurinic/apyrimidinic endonuclease (APEXL2 / APE2) | AP endonuclease | Abasic site |
| DNA polymerase β (Pol β) | DNA polymerase | Removal of 5'-terminal 2'-deoxyribose 5'-phosphate residue and incorporation of nucleotides |
| DNA polymerase δ or ε (Polδ or ε) | DNA polymerase | PCNA-dependent pathway |
| Flap endonuclease I (DNase IV) (FEN 1) | 5'-flap endonuclease | Removal of displaced region during PCNA-dependent pathway |
| Proliferating cell nuclear antigen (PCNA) | Forming a sliding clamp on DNA | Accessory factor for Pol δ/ ε and FEN1 |
| X-ray repair cross-complementing factor-1 (XRCC1) | Possible scaffold | Partner of Pol β, DNA ligase III |
| DNA ligase I (Lig I) | DNA ligase | Short and Long patch pathway |
| DNA ligase III (Lig III) | DNA ligase | Single nucleotide insertion pathway |

For example, thymine that is mismatched with guanine is recognized by thymine DNA glycosylase (TDG) and 8-oxo-7,8-dihydroguanine opposite cytosine, is recognized by human 8-oxoG DNA glycosylase (hOGG1). There are two types of glycosylases, monofunctional glycosylases (e.g. TDG, and uracil DNA glycosylase; UDG) and bifunctional glycosylase/AP lyases (e.g. hOGG1, and thymine glycol DNA glycosylase-AP lyase; hNth1). The DNA glycosylase cleaves the N-glycosydic bond that links the modified base with the sugar-

phosphate backbone by flipping out the damaged base into a binding pocket. The generated apyrimidic/apurinic (AP) sites are further incised on the 5'-side by an AP endonuclease (e.g. human apurinic/apyrimidinic endonuclease 1; APEX1) or on the 3'-side by an AP lyase. The 5'-terminal 2'-deoxyribose 5'-phosphate residue is removed and the correct nucleotide is incorporated into the gap by polymerase β (pol β). With monofunctional glycosylases, the repair can then follow two possible DNA synthesis pathways; a short patch (a single nucleotide) and a long patch (2-10 nucleotides) synthesises. On the other hand, bifunctional glycosylase prefers a single nucleotide replacement. In the short patch pathway, DNA is mainly synthesized by pol β, and ligated by DNA ligase I or by DNA ligase III in a complex with X-ray repair cross-complementing factor-1 (XRCC1). In the long patch pathway, a proliferating cell nuclear antigen (PCNA)-dependent polymerase δ/ε and pol β are involved in the synthesis step, and the gap in the phosphate backbone is sealed by DNA ligase I (Figure 10) (Frosina, 2000; Lindahl, 2000; Seeberg *et al.*, 1995). On the other hand, the bifunctional glycosylases follow the short-patch single nucleotide replacement pathway. There is no known catalytic function for XRCC1, however, it may play a role in the assembly of a repair complex since XRCC1 interacts with DNA pol β (Kubota *et al.*, 1996), and DNA ligase III (Cappelli *et al.*, 1997; Nash *et al.*, 1997). Polβ also shows interaction with APEX1 (Bennett *et al.*, 1997), and with DNA ligase I (Dimitriadis *et al.*, 1998; Prasad *et al.*, 1996).

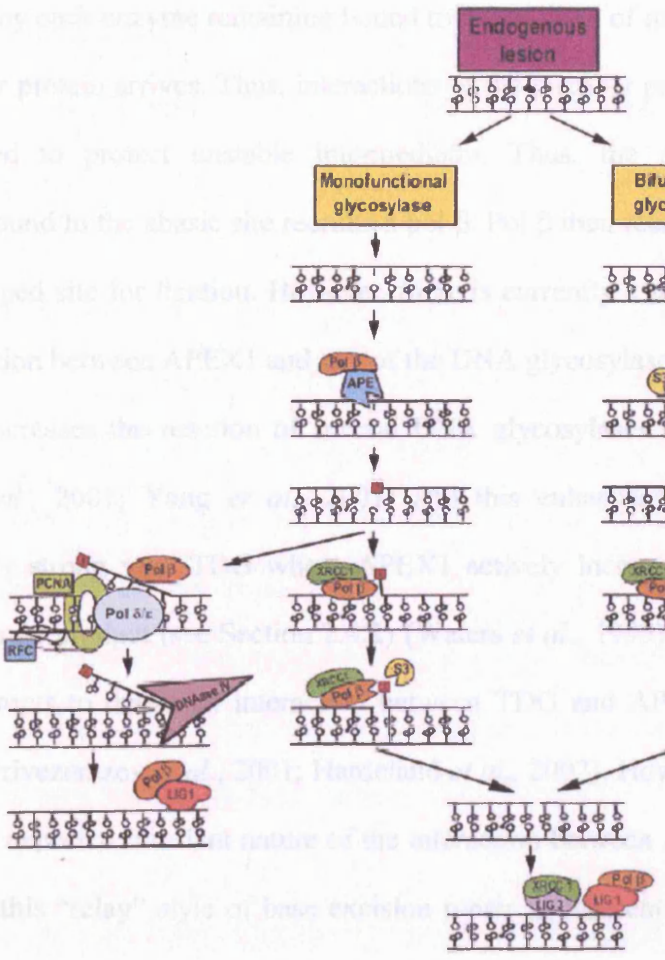



**Figure 10 Schematic steps of base excision repair in mammals**

**A glycosylase detects the damaged base and removes it, generating an AP site. In the short patch repair of the monofunctional glycosylase pathway, APEX1 binds to the AP site to nick the phosphodiester bond, followed by removal of abasic sugar-phosphate residue and incorporation of a nucleotide by Pol β. Pol β recruits XRCC1 in complex with Ligase III, or DNA ligase I, to seal the gap. This figure was taken from Frosina (2000).**

It is now apparent that each member protein of this repair system arrives at the damaged site when the previous reaction is completed. Therefore, the repair proceeds in a "relay" style with the DNA being handed from one enzyme to the next (Wilson and Kunkel, 2000). The intermediate at the damaged site is

protected by each enzyme remaining bound to the product of its reaction until the next repair protein arrives. Thus, interactions of the member proteins in BER are coordinated to protect unstable intermediates. Thus, the AP endonuclease, APEX1 bound to the abasic site recruits a pol β. Pol β then recruits DNA ligase I to the gapped site for ligation. However, there is currently a lack of evidence of an interaction between APEX1 and any of the DNA glycosylases. It is known that APEX1 increases the reaction of several DNA glycosylases (Hill *et al.*, 2001; Vidal *et al.*, 2001; Yang *et al.*, 2001) and this enhancement of activity is particularly strong with TDG where APEX1 actively increases dissociation of TDG from its product (see Section 2.4.2) (Waters *et al.*, 1999). There have been some attempts to detect an interaction between TDG and APEX1, but without success (Privezentzev *et al.*, 2001; Hardeland *et al.*, 2002). However, this may be due to the expected transient nature of the interaction between APEX1 and TDG. Note that this "relay" style of base excision repair is different from other repair pathways such as nucleotide excision repair (Section 2.3.3) and mismatch repair (Section 2.3.4) that form a multiprotein complex at the damage sites for repair.

## 2.3.3. Nucleotide Excision Repair

Bulky and helix distorting DNA damage, such as pyrimidine dimers caused by UV light, are repaired by nucleotide excision repair (Friedberg, 1985) (Figure 11).

Defects in nucleotide excision repair are linked to the inherited syndromes xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (Lehmann, 1995; Wood, 1997). The XP patients are separated into seven genetic complementation groups from XP-A to XP-G (corresponding to defects in each gene of the XPA to XPG proteins), and a variant type, XP-V. These syndromes give rise to sun-sensitive symptoms, leading to UV-induced skin cancers (XP), or severe developmental and neurological abnormalities (CS) (Lehmann, 1995; Wood, 1997).

Nucleotide excision repair involves the products of around 30 genes. These proteins remove the lesion by cutting out a short oligonucleotide (~32 nucleotides in eukaryotes). As a minimum model of mammalian nucleotide excision repair, a complex of XPA, RPA (a single stranded DNA binding protein) and XPC-hHR23B detects the lesion. The TFIIH complex containing the DNA helicases XPB and XPD is recruited to the lesion to open the DNA helix. A complex of ERCC1-XPF (a structure-specific nuclease) cleaves near the border between single-stranded and double-stranded DNA at the 5'-side of the lesion and the DNA endonuclease XPG cleaves at the 3'-side. The gap generated by the incision is filled by PCNA-dependent polymerase δ or ε holoenzyme. Then, the nick is sealed by a DNA ligase (Friedberg, 1996; Lehmann, 1995; Wood, 1997).

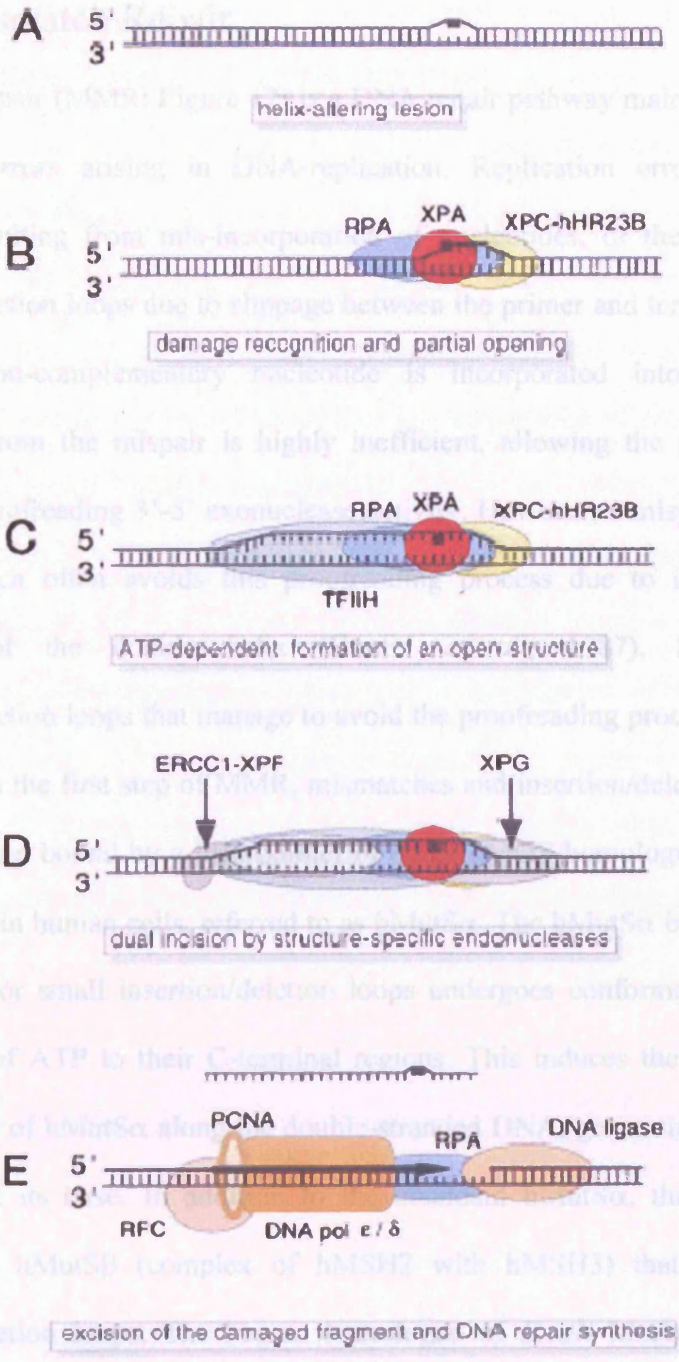



**Figure 11 Schematic model of nucleotide excision repair steps**

**A; damage of DNA, B; recognition of the damage, C; opening of the DNA double helix around the damage, D; cleavage of the damage site, E; repair synthesis. This figure was taken from Wood (1997).**

## 2.3.4. Mismatch Repair

Mismatch repair (MMR: Figure 12) is a DNA repair pathway mainly involved in correcting errors arising in DNA-replication. Replication errors are either mispairs resulting from mis-incorporation of nucleotides, or the formation of insertion/deletion loops due to slippage between the primer and template strands. When a non-complementary nucleotide is incorporated into the primer, elongation from the mispair is highly inefficient, allowing the polymerase to switch to proofreading 3'-5' exonuclease activity. However, a mispair, such as a G·T mismatch often avoids this proofreading process due to its only slight distortion of the double helix (Hunter *et al.*, 1987). Mispairs and insertion/deletion loops that manage to avoid the proofreading process are targets for MMR. In the first step of MMR, mismatches and insertion/deletion loops are recognized and bound by a heterodimer of human MutS homologue 2 (hMSH2) and hMSH6 in human cells, referred to as hMutSα. The hMutSα bound to single mismatches or small insertion/deletion loops undergoes conformational change by binding of ATP to their C-terminal regions. This induces the bi-directional translocation of hMutSα along the double-stranded DNA, generating a loop with the dimer at its base. In addition to the abundant hMutSα, there is another heterodimer, hMutSβ (complex of hMSH2 with hMSH3) that only detects insertion/deletion loops. The human homologue of *E.coli* MutL, hMutLα (or hMutLβ), is a heterodimer of hMLH1 and hPMS2 (or hMLH1/hMLH3) that interacts with hMutSα (or hMutSβ), enhancing the efficiency of hMutSα binding. Also, hMSH2 and hMLH1 are required to assemble the MMR repairsome (Gu *et al.*, 1998). In the MMR pathway, correction of bases is always

carried out in the daughter strand, which means that the system has to distinguish between template and the newly synthesized strands. Although discrimination of the strands by methylation status is used in *E.coli,* methylation patterns of vertebrates are irregular. For example, the CpG islands (up to several kilobases) are not generally methylated. Thus, it is unlikely that the vertebrate replication machinery uses methylation status to discriminate the two strands. Co-immunoprecipitation of hMLH1 showed an interaction with PCNA in addition to an interaction with hPMS2 and hMSH2 (Gu *et al.*, 1998). This evidence suggested that PCNA bound to the primer termini at the replication fork might be a discrimination signal (Umar *et al.*, 1996) via physical contacts with mismatch repair proteins. This interaction could lead to dissociation of the replication complex, allowing 3'-5' exonucleases to degrade the daughter strand (which contains the error). After dissociation of the MMR complex, PCNA could recruit the replication complex back again to resume DNA synthesis (Jiricny, 1998).

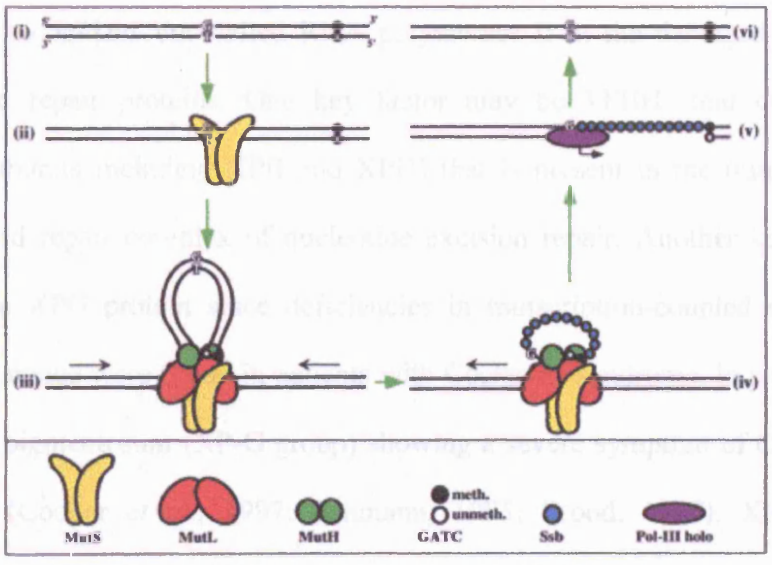



**Figure 12 Schematic model of mismatch repair in E.coli.**

**(i)A G·T mispair produced by an error of DNA polymerase. (ii) Recognition of the mismatch by MutS (iii) Bi-directional translocation of the mispair by a multi-protein complex, forming a loop. MutH (an endonuclease) cleaves the daughter strand. (iv) The daughter strand is degraded from the nick to the mismatch site by ExoVII or RecJ. The single-stranded part is covered by single strand-binding protein (Ssb). (v) Polymerase III holoenzyme fills the gap, then the nick is sealed by DNA ligase. This figure was taken from Jiricny (1998).**

## 2.3.5. Transcription-coupled repair

It is known that certain DNA damage, such as that induced by UV light or oxidation (e.g. thymine glycols), in the transcribed strand of a gene is repaired prior to the non-transcribed strand. This pathway is called transcription-coupled repair (TCR). TCR requires gene products defective in Cockayne syndrome. However, this strand-specific repair of transcribed genes is limited to genes transcribed by the RNA polymerase II complex. In order to couple nucleotide excision repair to TCR, and base excision repair to TCR there must be a common

factor that recognizes the stalled RNA polymerase II at the damaged site and recruits the repair proteins. One key factor may be TFIIH (that comprises multiple subunits including XPB and XPD) that is present in the transcription complex and repair complex of nucleotide excision repair. Another key factor may be the XPG protein since deficiencies in transcription-coupled repair of oxidative damage were found in patients with Cockayne syndrome, in patients of xeroderma pigmentosum (XP-G group) showing a severe symptom of Cockayne syndrome (Cooper *et al.*, 1997; Lehmann, 1995; Wood, 1997). XPG from nucleotide excision repair is known to help load hNTH (a bifunctional BER glycosylase that removes thymine glycol) onto the damaged site (Klungland *et al.*, 1999; Lindahl, 2000), indicating the sharing of common proteins between TCR, nucleotide excision repair and base excision repair (Cooper *et al.*, 1997). Therefore, it seems that XPG has two different functions; an endonuclease in nucleotide excision repair and an assembly factor in base excision repair.

There is an implied relationship between mismatch repair and nucleotide excision repair pathways. Studies of *E.coli* cells that had mutant mutS or mutL mismatch repair genes indicated a role of these proteins in TCR since they lack the strand selective repair of UV damage (Lahue, 1996).

There is also the possibility of involvement of other proteins in TCR of oxidative DNA damage. BRCA1 or BRCA2 (the breast and ovarian cancer susceptible genes) are likely to participate in TCR since deficiency of either gene causes the transcriptional machinery to pause at 8-oxoguanine (Page *et al.*, 2000). Thus, the TCR system might interconnect mismatch repair, nucleotide excision repair, and base excision repair pathways by sharing common factors to restore genetic information where repair is urgently required for smooth transcription.

## 2.3.6. Repair of Double-strand break

Reactive oxygen species and ionizing radiation can break the DNA backbone. Eukaryotes have developed two pathways to repair double-strand breaks: homologous recombination and nonhomologous end joining.

Homologous recombination plays the role when a sister chromatid is available as a template in the S and G2 phases of the cell cycle. Homologous recombination involves many proteins, and the pathway is complex. Briefly, the end of the break is processed to make a 3' single-stranded overhang tail. The tail is used to search for a homologous template. When the 3' tails form a duplex with the homologous template, DNA polymerase incorporates nucleotides between the gaps, and the nicks are sealed. The structure of branched DNA is now called Holliday junctions. It migrates along the joined DNA and is resolved to form two copies of duplex DNA (Figure 13).

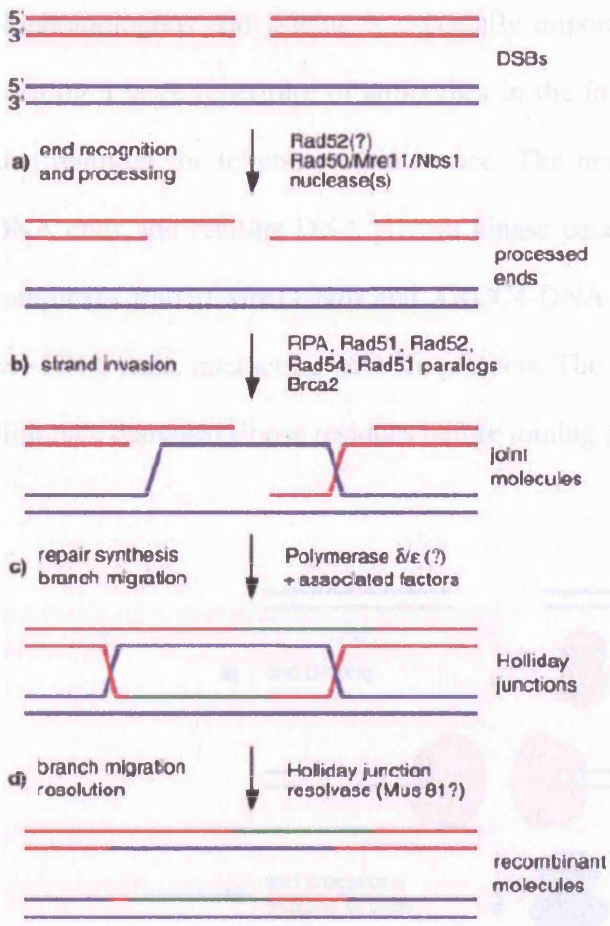



**Figure 13 Schematic diagram of repair of double-strand break by homologous recombination**

**a) The damaged DNA (red) may be recognized by Rad52 then the complex of Rad50/Mre11/Nbs1 may generate 3'-ssDNA overhangs. b) The ssDNA ends are bound by RPA with help of other proteins and Rad52 is loaded onto the ssDNA. The ssDNA bound by Rad52 searches for homologous duplex DNA (blue), and a strand-exchange reaction occurs. c) DNA polymerases and their associated factors fill the gaps and Holliday junctions are formed. d) Holliday junctions are resolved by endonucleolytic cleavage and rejoined.**

**The figure was taken from Schärer (2003).**

Nonhomologous end joining is especially important to join DNA fragments in creating a large repertoire of antibodies in the immune system. This pathway is also required for telomere maintenance. The heterodimer Ku70/Ku80 binds to DNA ends and recruits DNA protein kinase catalytic subunit (DNA-$PK_{cs}$). The complexes Rad50-Mre11-Nbs and XRCC4-DNA ligase IV come in between the two DNA ends, interacting with Ku proteins. The DNA ends may be processed to eliminate damaged ribose residues before joining (Figure 14) (Schärer, 2003).

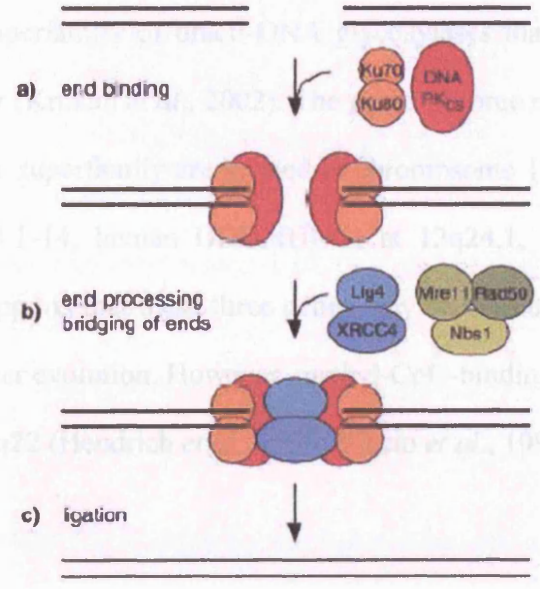



**Figure 14 Schematic diagram of double-strand breaks repaired by nonhomologous end joining**

**a) Double-strand breaks are recognized by the Ku70 and Ku80, which specifically bind to DNA ends and form a complex with the DNA-dependent protein kinase catalytic subunit (DNA-$PK_{cs}$). b) The ends of the break may be processed with Rad50/Mre11/Nbs1 and a further nuclease. The Rad50/Mre11/Nbs1 and Xrcc4/DNA ligase IV complexes might be bridging the DNA ends. c) The two ends are ligated by xrcc4/DNA ligase IV. The figure was taken from Schärer (2003).**

# 2.4. Uracil-DNA glycosylases

## 2.4.1. Human Uracil-DNA glycosylase

In humans there are four glycosylases that can remove uracil from DNA; the general uracil-DNA glycosylase (UDG), single strand specific uracil-DNA glycosylase (SMUG), TDG and methyl CpG binding domain protein 4 (MBD4). Despite low primary sequence similarity between these enzymes, structural studies show that the first three uracil glycosylases (UDG, SMUG, and TDG) have very similar structures. It is now believed that these three glycosylases are members of a superfamily of uracil-DNA glycosylases that are derived from a common ancestor (Krokan *et al.*, 2002). The genes of three members of the uracil DNA glycosylase superfamily are located in chromosome 12; human SMUG1 is located at 12q13.1-14, human UDG (UNG) at 12q24.1, and human TDG at 12q24.1. This supports that these three genes may be duplicated from a common ancestral gene over evolution. However, methyl-CpG-binding domain protein 4 is located on 3q21-q22 (Hendrich *et al.*, 1999; Riccio *et al.*, 1999).

*Uracil-DNA glycosylase*

*E.coli* uracil-DNA glycosylase was first discovered by Tomas Lindahl (1974). UDG is highly and widely conserved among organisms from herpes simplex virus type I, *E.coli* to human (more than 40 % of amino acid sequence identity). Human UDG can cleave the uracil derivatives isodialuric acid, 5-hydroxyuracil, and alloxan, and products of cytosine oxidation as well as uracil (Dizdaroglu *et al.*, 1996). The crystal structure of UDG was the first DNA glycosylase to be

solved and it revealed that the glycosylase flips the uracil base out of the DNA helix into a binding pocket on the protein where the cleavage of the glycosidic bond takes place.

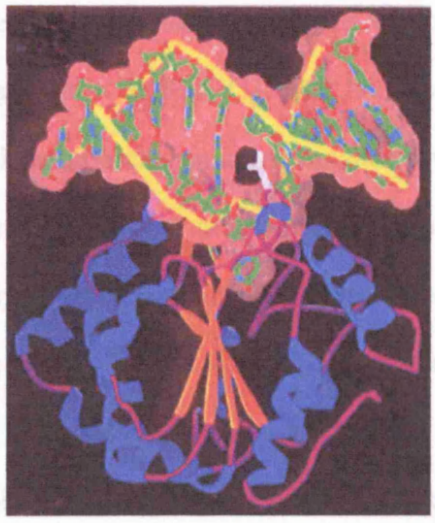


**Figure 15 Crystal structure of human uracil-DNA glycosylase complexed with DNA containing U·G mismatch**

**This figure was taken from Parikh *et al.* (1998).**

UDG has a highly specific binding pocket and can remove uracil from single- and double-stranded DNA (Delort *et al.*, 1985). Formation of uracil in DNA by spontaneous deamination of cytosine is one of the most common type of DNA damage, and it is estimated to occur 60-500 times per genome per day (Lindahl, 1993). Possibly in order to cope with this high frequency of damage, UDG has developed as the most efficient glycosylase so far discovered. Turnover of UDG is not inhibited by its product, and the $k_{cat}$ of UDG from herpes simplex virus type I is approximately 6 $s^{-1}$ for removal of uracil from single-stranded substrate and 2 $s^{-1}$ for the removal of uracil from double-stranded substrate (Bellamy and

Baldwin, 2001). In addition to uracil arising from deamination of cytosine, dUMP is sometimes mistakenly incorporated into DNA during replication. The resultant A·U base pairs are also substrates for UDG. UDG knockout mice do not display significantly elevated spontaneous mutation rate and have no obvious pathological phenotype (Nilsen *et al.*, 2000). The substantial backup of uracil excision activity was attributed to SMUG1. The knockout mice were found to have slow repair of misincorporated uracil mismatched with adenine. This fact together with the diminutive increase of spontaneous mutation rate imply that a major role of UDG may be repair of misincorporated uracil rather than repair of deaminated cytosine *in vivo* (Nilsen *et al.*, 2000). In humans there are two major splice forms of UDG with different promoters; UNG1 (found mainly in the mitochondria) and UNG2 (found mainly in the nucleus) (Haug *et al.*, 1998; Nilsen *et al.*, 1997) and it appears that a fraction of the nuclear UDG is associated with DNA replication forks in order to repair uracil that has been misincorporated during replication (Nilsen *et al.*, 2000; Otterlei *et al.*, 1999). Interestingly, it is reported that uracil-DNA glycosylase may be involved in a pathway that gives rise to diversification of B-cell antibodies. In activated B cells, activation-induced deaminase is specifically expressed. The protein deaminates deoxycytidine residues in the immunoglobulin loci (Petersen-Mahrt *et al.*, 2002). Therefore, a substrate for UDG is generated by enzymatic deamination of cytosine, and UDG may play a part in gene diversification (Neuberger *et al.*, 2003). UDG deficient mice have significantly inhibited class-switch recombination (switching from IgM to IgG/A/E), which involves KU70/KU80 of nonhomologous recombination repair (Rada *et al.*, 2002).

0

*SMUG1 (Single strand-selective Monofunctional Uracil-DNA Glycosylase)*

Human SMUG1 contains approximately 270 amino acids (30 kDa). SMUG1 removes uracil and 5-hydroxymethyluracil from single- and double-stranded DNA. However, SMUG1 excises uracil from a double-stranded DNA more efficiently than from a single-stranded DNA by nearly 700-fold (Wibley *et al.*, 2003). It now seems likely that in humans UDG is involved in post-replicative repair of uracil from A·U base pairs whereas SMUG1 is responsible for removing uracil generated by deamination of cytosine in bases pairs with guanine (Nilsen *et al.*, 2001). SMUG1 inefficiently excises ethenocytosine (Kavli *et al.*, 2002), but does not cleave thymine mismatched with guanine (Hardeland *et al.*, 2001). The apparent $k_{cat}$ of SMUG1 against uracil mispaired with guanine and adenine are 9.1 $s^{-1}$ and 0.36 $s^{-1}$, respectively (Wibley *et al.*, 2003).

*Methyl-CpG-binding domain protein 4*

A second protein that excises thymine mismatched with guanine was discovered by Hendrich and Bird (1998) (Hendrich and Bird, 1998). The protein is called MBD4 (methyl-CpG-binding domain protein 4) or MED1 (methyl-CpG-binding endonuclease) containing approximately 580 amino acids. MBD4 contains an amino-terminal methyl-CpG-binding domain and a carboxy-terminal glycosylase domain (amino acids 379-580) that exhibits homology to *E.coli* endonuclease III and MutY (Bellacosa *et al.*, 1999). TDG does not contain a methyl-CpG-binding domain and does not share amino acid sequence similarity with MBD4 even though they show quite similar substrate specificity. MBD4 is able to remove uracil in a G·U mismatch faster than thymine in a G·T mismatch, but it cannot

process C·A, C·C or G·G mismatches (Hendrich *et al.*, 1999). It also excises the uracil analogue, 5-fluorouracil from mismatches with guanine (Petronzelli *et al.*, 2000), 3,$N^4$-ethenocytosine (Petronzelli *et al.*, 2000), and thymine glycol at half the rate of excision of thymine (Yoon *et al.*, 2003). MBD4, like TDG, prefers a CpG sequence for the mismatches, and shows preference for $^{5\text{-Me}}$CpG·T rather than CpG·T and a strong binding to the AP site ($k_{off} \approx 8 \times 10^{-6}$ $s^{-1}$; Petronezelli *et al.*, 2000). Interestingly, excision of 5-methylcytosine from hemi-methylated DNA was been reported as an activity of MBD4 (Zhu *et al.*, 2000). Furthermore, MBD4 may have a role in mismatch repair since an interaction with MLH1 has been detected (Bellacosa *et al.*, 1999), and it may be involve in the apoptotic response to damaging agents in which mismatch repair proteins have been shown to be involved (Bellacosa, 2001; Bellacosa *et al.*, 1999; Buermeyer *et al.*, 1999). It is suggested that MBD4 might be a eukaryotic homologue of the bacterial MMR endonuclease, MutH (Bellacosa *et al.*, 1999). Unlike TDG (Schmutte *et al.*, 1997), mutation of MBD4 is reported to be involved in human cancer (Riccio *et al.*, 1999), especially related to microsatellite instability tumours (Bader *et al.*, 1999; Boland *et al.*, 1998; Dietmaier *et al.*, 1997).

## 2.4.2. Thymine-DNA glycosylase

Thymine-DNA glycosylase (TDG) was discovered from extracts of HeLa cells as an enzyme that cleaves the glycosydic bond of thymine mismatched with guanine by Wiebauer *et al.* in 1987 (Wiebauer and Jiricny, 1989). Karan and Griffin (1993) reported that a G·T mismatch binding protein from extracts of mammalian cells showed most efficient excision of thymine when the G·T mismatch was in a CpG or a $^{5\text{-me}}$CpG sequence while the substrates GpG·T, ApG·T, and TpG·T were not efficiently incised (Griffin and Karran, 1993). Subsequently, Neddermann *et al.* (1996) cloned the cDNA of human thymine DNA glycosylase, and reported that the mismatch-specific glycosylase contains 410 amino acids (see Appendices), and has a mass of 46 kDa (although it migrates around the 60 kDa-position in denaturing polyacrylamide gels; Neddermann *et al.*, 1996).

## 2.4.3. Characteristics of thymine-DNA glycosylase

The gene of human thymine-DNA glycosylase was mapped to 12q24.1 of chromosome 12 by Sard *et al.* (1997) and Schmutte *et al.* (1997). The level of mRNA of TDG has been reported to be constant, in other words, expression of TDG is not cell cycle-dependent (Bouziane *et al.*, 2000). Niederreither *et al.* (1998) reported that TDG was equally and ubiquitously expressed until ~13.5 days of the developing mouse foetus, but that the level of expression was significantly enhanced from 14.5 days, which may be correlated with high expression of the methyl transferase that methylates cytosine in CpG

dinucleotides. The areas of high expression of TDG were generally allocated to the proliferating cells of different organs. They also reported that TDG and the methyl transferase were strongly expressed in experimentally induced mouse tumours. However, the implication of the high expression of TDG in the tumours is not clear (Niederreither *et al.*, 1998). Also, the possibility of induced expression of TDG in response to DNA damage cannot be dismissed.

*Substrate Specificity*

TDG shows the following preference for DNA binding; G·AP site > G·U > G·T >> C·T > T·T (Schärer *et al.*, 1997; Sibghat-Ullah *et al.*, 1996). Consistent with this, TDG also shows a very strong preference for excision of thymine mismatched with guanine (Sibghat-Ullah *et al.*, 1996). The rate of thymine excision is also very dependent upon the sequence surrounding the G·T mismatch (Griffin *et al.*, 1994; Griffin and Karran, 1993; Schärer *et al.*, 1997; Sibghat-Ullah and Day, 1995; Sibghat-Ullah *et al.*, 1996; Waters and Swann, 1998). A study from this laboratory measuring single-turnover reaction rates for thymine excision found the following order of sequence preference; CpG·T ($k_{cat}$ = 0.91 $min^{-1}$) >> TpG·T ($k_{cat}$ = 0.023 $min^{-1}$) > GpG·T ($k_{cat}$ = 0.0046 $min^{-1}$) > ApG·T ($k_{cat}$ = 0.0013 $min^{-1}$) (Waters and Swann, 1998). Therefore, the sequence CpG·T is the most favoured substrate, being at least 40 times faster than other sequences. TDG therefore recognises both the G·T mismatch itself and the neighbouring G·C base pair 5' to the mismatched guanine. Crystal structures of the bacterial homologue of TDG show that the preference for mismatches with guanine is probably due to TDG making hydrogen bond contact with the mismatched guanine (Barrett *et al.*,

1998; Barrett *et al.*, 1999). Also, methylation interference experiments by Schärer *et al.* (1997) confirmed that TDG makes contact with guanine residue of the neighbouring G·C base pair 5' to the mismatch (Schärer *et al.*, 1997). The strong preference for G·T mismatches at CpG sequences supports the hypothesis that TDG has evolved to cope with deamination of 5-methylcytosine, since 5-methylcytosine always occurs at CpG sites and so deamination of 5-methylcytosine would always produce a CpG·T mismatch.

In addition to thymine, *in vivo* TDG can excise uracil (CpG·U; Apparent $k_{cat}$ = 11 $min^{-1}$, CpC·U; $k_{cat}$ = 4.7 $min^{-1}$, CpT·U; $k_{cat}$ = 1.2 $min^{-1}$, and CpA·U; $k_{cat}$ = 0.047 $min^{-1}$) (Neddermann and Jiricny, 1994; Waters and Swann, 1998), ethenocytosine (εC) (see Figure 16)(Hang *et al.*, 1998; Saparbaev and Laval, 1998), 5-fluorouracil, 5-hydroxyuracil, and 5-hydroxymethyluracil (Hardeland *et al.*, 2001) from double-stranded DNA (Krokan *et al.*, 2002).

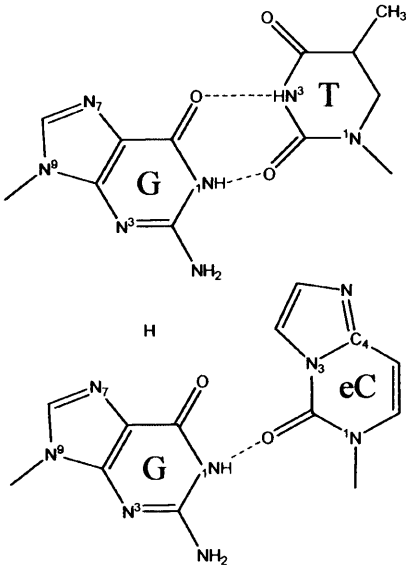



**Figure 16 Schematic diagrams of G·T and G·εC mismatches**

**The dotted line indicates possible hydrogen bond between the base pairs. The diagram of ethenocytosine mismatched with guanine was derived from Cullinan *et al.* (1997), and thymine mismatched with guanine from Hunter *et al.* (1987) and Allawi and SantaLucia (1998).**

*Tight binding of thymine-DNA glycosylase to the product*

In the base excision repair of G·T mismatches, TDG first excises the damaged base, but then remains tightly bound to the apurinic/apyrimidinic site product. That is, TDG catalyses in a stoichemetric way, since in the absence of a releasing factor, each molecule of TDG can remove only one molecule of thymine. It has been suggested that TDG stays on the product to protect the unstable intermediate until the next protein, apurinic/apyrimidinic endonuclease 1 (APEX1) arrives at the site. The rate of dissociation of TDG from the abasic site can be increased in a concentration dependent manner by addition of APEX1 (Figure 17) (Waters *et al.*, 1999). Therefore, it is thought that there may be a protein-protein interaction between TDG and APEX1, and that this interaction would induce dissociation of TDG from the abasic site.

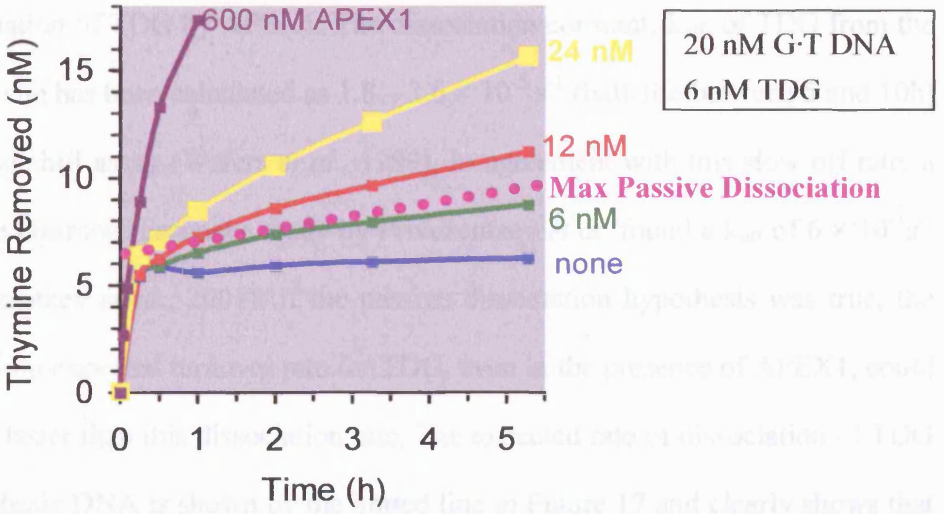



**Figure 17 Dissociation rate of TDG in the presence of APEX1 indicating the maximum passive dissociation**

**The curve for the maximum passive dissociation was calculated using a $k_{off}$ of $2.8 \times 10^{-5}$ $s^{-1}$, derived from a bandshift experiment (Waters et al., 1999). 6 nM of TDG, 20 nM of oligonucleotide containing a G·T mismatch, and none or 6, 12, 24, and 600 nM of APEX1 were incubated, at room temperature. As the concentration of APEX1 increases, TDG is dissociated more rapidly from the abasic site and allowed to catalyse further reaction. In the absence of APEX1, TDG could only catalyse a stoichemetric amount of substrate. The data was taken from Waters *et al.* (1999).**

However, another hypothesis has been suggested to explain the increased turnover of TDG in the presence of APEX1. In this hypothesis, TDG dissociates slowly from the AP site but in the absence of APEX1 it rebinds to the abasic site rather than bind, and react with, another G·T mismatch. In the presence of APEX1, TDG and APEX1 simply compete with each other for the abasic site and so increasing concentrations of APEX1 prevent the TDG from re-binding abasic DNA and force the glycosylase to bind to further GT mismatches. Therefore, in

this 'passive' model there is no protein-protein interaction and no active dissociation of TDG by APEX1. The dissociation constant, $k_{off}$, of TDG from the abasic site has been calculated as $1.8 - 3.6 \times 10^{-5}$ $s^{-1}$ (half-life between 5 and 10h) by band shift assay (Waters *et al.*, 1999). In agreement with this slow off rate, a surface plasmon resonance study by Privezentzev *et al.* found a $k_{off}$ of $6 \times 10^{-5}s^{-1}$ (Privezentzev *et al.*, 2001). If the passive dissociation hypothesis was true, the maximum expected turnover rate for TDG, even in the presence of APEX1, could not be faster than this dissociation rate. The expected rate of dissociation of TDG from abasic DNA is shown by the dotted line in Figure 17 and clearly shows that APEX1 increases the turnover of TDG to a rate far higher than the expected rate of passive TDG dissociation.

The passive model therefore does not explain the accelerated rate of TDG reaction in the presence of APEX1 in Figure 17. In addition, the displacement of TDG is specific to APEX1 since the *E.coli* AP endonucleases Formamidopyrimidine-DNA glycosylase, Endonuclease IV, and Endonuclease III, do not displace TDG from the abasic site (Privezentzev *et al.*, 2001; Waters *et al.*, 1999). These experiment are very interesting since they showed that the abasic site was not available for competition, meaning that TDG protected the abasic site from access by the AP endonucleases. Thus, this is further evidence that TDG does not dissociate from the abasic site on its own, and APEX1 may specifically dissociate TDG via a direct protein-protein interaction. The yeast two-hybrid system was employed by Privezentzev *et al.* (2001) to try and detect protein-protein interactions between TDG and APEX1, but they found no evidence of such an interaction. (Privezentzev *et al.*, 2001). However, the protein-protein interaction may be extremely transient and would therefore be very

difficult to detect (Tini *et al.*, 2002). Base excision by other glycosylases can also be stimulated by APEX1. Vidal et al. (2001) found that the activity of hOGG1 was stimulated in the presence of APEX1, however, this acceleration was probably due to passive competition for abasic DNA since the dissociation constant of hOGG1 from the AP site was not decreased by APEX1 (Vidal *et al.*, 2001). The activity of hOGG1 could also be stimulated by the bacterial AP endonuclease IV (product of nfo gene) even though this endonuclease is not related to APEX1 (Hill *et al.*, 2001; Vidal *et al.*, 2001). This suggests that the stimulation of hOGG1 by those endonucleases might not be specific, and that there is a fundamental difference in the mechanism of dissociation between hOGG1 and TDG.

*Crystal Structure of the Bacterial Homologue of TDG*

TDG has an *E coli.* homologue, called mismatch-specific uracil DNA glycosylase (MUG). The amino acid sequence of MUG is 120 and 100 residues shorter than TDG at the amino terminus and the carboxyl terminus, respectively. However, the core part of the enzyme (residues 112-360 of TDG) is conserved more than 30 % between MUG and TDG (Gallinari and Jiricny, 1996). MUG can excise uracil (Gallinari and Jiricny, 1996), ethenocytosine, thymine (Saparbaev and Laval, 1998), 8-(hydroxymethyl)-3,$N^4$-etheno- 2'-deoxycytidine (8-HM-εdC) (Hang *et al.*, 2002), 5-hydroxymethyluracil, 5-hydroxyuracil, inosine (hypoxanthine), and 1,$N^2$-ethenoadenine (O'Neill *et al.*, 2003). Ethenocytosine is the fastest substrate to be excised (O'Neill *et al.*, 2003). However, it may not be the real *in vivo* target of MUG since ethenocytosine has not yet been found in *E.coli* DNA (Bartsch and

Nair, 2000). MUG, like TDG, displays strong binding to its product, the abasic site. Also, like TDG, the activity of MUG is enhanced on addition of an AP endonuclease, *E.coli* endonuclease IV (Hang *et al.*, 2002). The crystal structure of the *E.coli* homologue of TDG, mismatch-specific uracil DNA glycosylase (MUG) gives us a good insight into the mechanism of excision by TDG (Figure 18 and Figure 19). MUG (and by implication TDG) flips out the damaged base into a binding pocket on the enzyme, where it is excised. The structurally similar uracil DNA glycosylase (UDG) has a pocket that is strictly specific for uracil (Figure 20). In the case of MUG (and TDG), double-stranded DNA is required for excision, implying that they need to interact with the second DNA strand in order to displace the damaged base from the DNA helix.

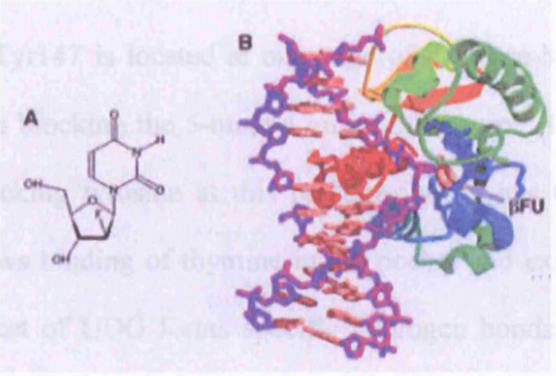
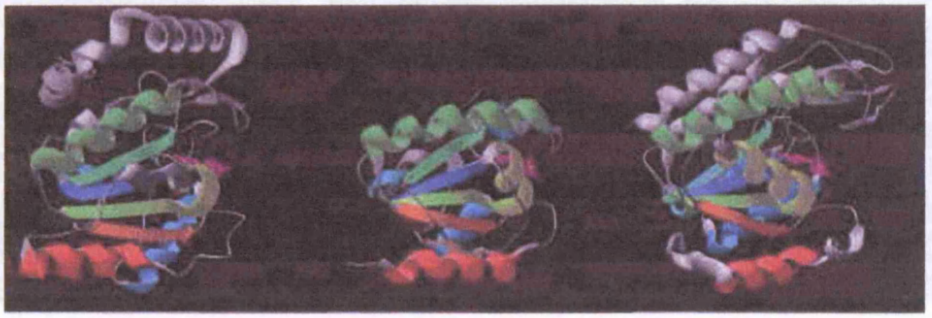



**Figure 18 Crystal structure of mismatch-specific uracil DNA glycosylase in complex with DNA**

**A; 1-(2'-deoxy-2'-fluoro-β-D-arabino furanosyl)–uracil B; MUG forms complex with DNA that contains an excision-resistant analogue of uracil. The figure was taken from Barrett *et al.* (1999).**

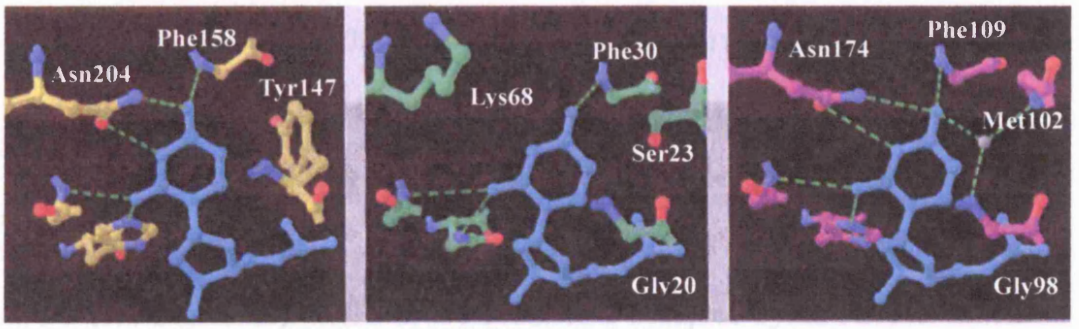


**Human UDG** ***E.coli* MUG** ***Xenopus laevis* SMUG1**

**Figure 19 Comparison of crystal structures of uracil-DNA glycosylase (human), mismatch-specific uracil DNA glycosylase (*E.coli*) and *Xenopus laevis* SMUG1**

**The secondary structure elements are colour coded to highlight the similarity in structure between the three glycosylases. The UDG structure from Mol *et al.* (1995), the MUG structure from Barrett *et al.* (1998) and the SMUG1 structure from Wibley *et al.* (2003)**

In human UDG, Tyr147 is located at one edge of the base-binding pocket and plays a key role in blocking the 5-methyl group of thymine (Figure 20). Instead of having the blocking tyrosine at this position, MUG has Gly20 (Gly142 in TDG), which allows binding of thymine in the pocket and excision of the base. Uracil in the pocket of UDG forms specific hydrogen bonds with a conserved Asn204; the amide side chain forms a hydrogen bond with the O4 carbonyl oxygen of uracil and the carbonyl oxygen accepts a hydrogen bond from the protonated N3 of uracil. Furthermore, the O4 of uracil forms another hydrogen bond with the peptide nitrogen of Phe101. The asparagine and the phenylalanine might expel the N4 of cytosine.




**Human UDG** ***E.coli* MUG (TDG)** ***Xenopus laevis* SMUG1**

**Figure 20 Comparison of binding pockets of human UDG, *E.coli* MUG and *Xenopus laevis* SMUG1**

**In human UDG, uracil has specific contacts with Asn204 and Phe158. Tyr147 blocks 5-methylgroup of thymine. *E.coli* MUG does not have the barrier and can therefore have thymine in the pocket. Structurally similar *Xenopus laevis* SMUG1 also expels thymine since 5-methylgroup of thymine is too large and the water molecule block the 5-methylgroup. The figure was derived from Mol *et al.* (1995), Barrett *et al.* (1998) and Wibley *et al.* (2003).**

In contrast, MUG cannot have the same specific interaction with uracil in the pocket since it has Lys68 (but TDG has Asn191) at this position (Barrett *et al.*, 1999). This means that the base-binding pocket of MUG cannot differentiate uracil and thymine from cytosine, and it is likely that MUG's catalytic specificity comes mainly from the ease of disruption of the inherent weak G·U base pair in comparison to the stronger G·C base pair. This lack of specific interaction in the pocket is illustrated by the fact that MUG can remove ethenocytosine when mismatched with guanine (O'Neill *et al.*, 2003). TDG allows thymine in its pocket since the 5-methyl group of thymine is not hindered by Gly142 (Tyr147 in hUDG), and may be able to differentiate uracil and thymine from cytosine due to

hydrogen bonds between O4 of thymine and side chain of Asn191 of TDG (equivalent to Asn204 in hUDG), but cannot differentiate thymine from uracil.

MUG shows a 14,800-fold faster cleavage rate for G·U over G·T (O'Neill *et al.*, 2003). Barrett *et al.* (1999) suggested that this might be caused by a moderate steric clash between the 5'-methyl of thymine and the side chain of Ser23 of MUG (see Figure 20). In TDG, the position is occupied by an Ala145 which would create less steric clash than serine and probably explains why TDG has only a 10-fold preference for uracil over thymine (Waters and Swann, 1998). One very interesting finding from the structure of MUG is that it (and TDG) has developed a mechanism for substrate specificity via contact with the second DNA strand. The residues Ala77 and Ser142 (Ser271 in TDG) of MUG form hydrogen bonds with the phosphate 3' to the abasic site. The phosphate 5' to the AP site makes contact with residues Ser22, Ser23, and Gly20 (Gly142).

The residues Gly143, Leu144, and Arg146 (Arg275 in TDG) are inserted between two DNA bases, into the space produced by the flipped out uracil. The loop is inserted from the minor groove, acting like a wedge that widens the space between the adjacent bases. The side chain of Arg146 forms two hydrogen bonds, one with the deoxyribose ring oxygen of the widowed guanine (i.e. the guanine from the G·U mismatch), and the other with the peptide oxygen of Leu144. On displacement of the deoxyuridine, three secure hydrogen bonds are formed with the widowed guanine; between the N1 imino group of the guanine and the carbonyl oxygen of Gly143, and between the N2 exocyclic amino group of the guanine and the carbonyl oxygen of Gly143 and the carbonyl of Ser145 (Figure 21) (Barrett *et al.*, 1998). These interactions are very important since they are

specific for guanine, and enable MUG to discriminate between G·U mismatches and A·U base pairs (and between G·T mismatches and A·T base pairs for TDG).

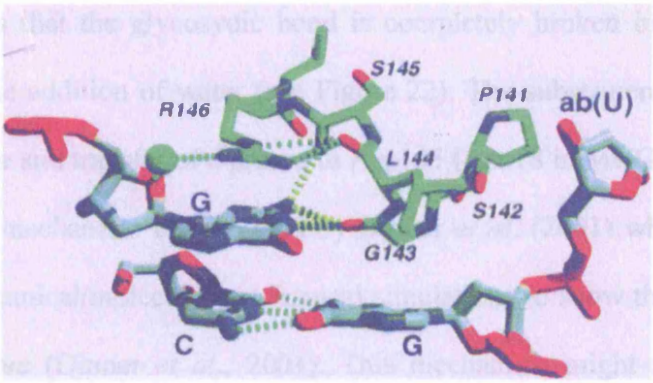



**Figure 21 Interaction between the widowed guanine and wedge (Gly143, Leu144, and Arg146) of mismatch-specific DNA glycosylase**

**The figure was taken from Barrett *et al.* (1998).**

Barrett *et al.* (1998) suggested that MUG uses a "push" mechanism to bind uracil in the pocket, rather than the "pull" mechanism employed by UDG (Kunkel and Wilson, 1996), which is promoted by inserting the wedge (Gly143, Leu144, and Arg146) into the DNA via the minor groove to flip out the uracil. Since MUG lacks the specific interaction between the flipped-out uracil and the pocket, it lacks the "pull" utilized by UDG (Barrett *et al.*, 1998).

*Catalytic Mechanism of Uracil-DNA glycosylase*

By measuring kinetic isotope effect on uracil excision by UDG, Werner *et al.* (2000) found that the reaction of UDG proceeds through a dissociative transition state, which is that the glycosydic bond is completely broken in the transition state before the addition of water (see Figure 22). The subsequent attack by the water molecule and transfer of a proton to Asp145 (Asn18 in MUG) generates the product. This mechanism is supported by Dinner *et al.* (2001) who used hybrid quantum-mechanical/molecular-mechanical simulations to show that the reaction was dissociative (Dinner *et al.*, 2001). This mechanism might apply to other glycosylases.

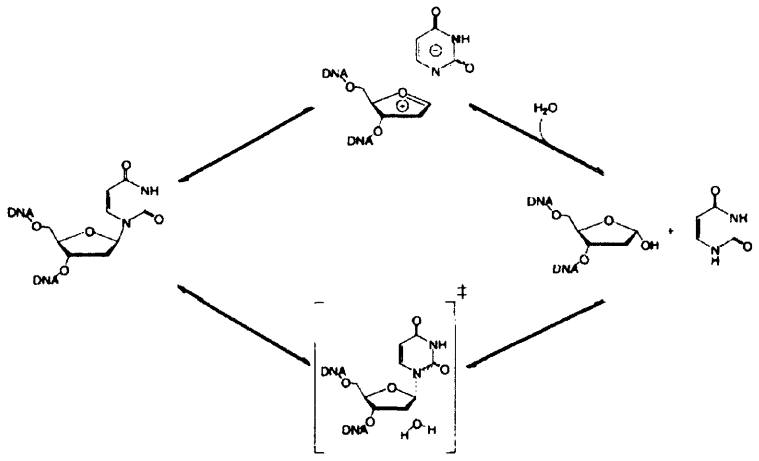



**Figure 22 Possible mechanisms for uracil-DNA glycosylase**

**The top pathway is a stepwise dissociative mechanism that produces a discrete oxocarbenium cation and a uracilate anion in the transition state followed by nucleophilic attack of a water molecule. The lower pathway illustrates the concerted associative mechanism that involves nucleophilic attack of a water molecule and expulsion of the uracil-leaving group in the transition state. The figure was taken from Werner and Strivers (2000).**

*Interaction of thymine-DNA glycosylase with other proteins*

TDG is reported to interact with several proteins involved in transcriptional activation. The retinoid acid receptor (RAR) and the retinoid X receptor (RXR) are transcription factors that are activated through ligand binding. Using a yeast one-hybrid assay TDG, was shown to interact with RAR/RXR and to enhance RAR/RXR dependent gene expression *in vitro*. Also, Chen *et al.* showed that TDG interacts with the oestrogen receptors and enhances oestrogen receptor mediated-gene regulation (Chen *et al.*, 2003). Furthermore, the thyroid transcription factor-1, which is involved in thyroid, lung, and ventral forebrain development (Kimura *et al.*, 1996; Sussel et al., 1999), also interacts with TDG (Missero *et al.*, 2001). However, in this case TDG suppresses thyroid transcription factor-1-activated transcription (Missero *et al.*, 2001). TDG was first isolated as a protein that could interact, via its C-terminal, with the leucine zipper of the transcription factor, c-Jun (Chevray and Nathans, 1992). Interestingly, c-Jun is activated by a redox reaction induced by Ref-1, a major nuclear redox factor that is identical to the BER enzyme APEX1. APEX1 is thus a bifunctional protein that has AP endonuclease activity, and redox activity that reduces the DNA binding domains of Fos and Jun to enhance DNA-binding activity (Kelley *et al.*, 2001; Xanthoudakis *et al.*, 1992). This suggests an interesting relationship between TDG and APEX1 via Fos/Jun that seems to have interactions with both TDG and APEX1. Several cancer cells have elevated levels of APEX1 indicating promoted activity of Fos/Jun proteins in these cells, and this may affect TDG activity (Kelley *et al.*, 2001). In addition to directly interacting with these transcription factors, TDG also interacts with certain coactivators. CREB binding protein (CBP) (Chrivia *et al.*, 1993) and p300 (Eckner *et al.*, 1994) act as

coactivators for many sequence-specific transcription factors, including nuclear receptors, Jun/Fos, and p53. Probably, CBP/p300 activates transcription via its intrinsic histone acetyltransferase activity that opens up the chromatin structure. Tini *et al.* (2002) found that TDG interacts with, and is acetylated by, CBP/p300, and that this interaction stimulates CBP related-transcription. In addition, they showed that mouse TDG interacts with APEX1, but that acetylation of TDG by CBP/p300 prevents this interaction with APEX1. They suggested that acetylation might work as a molecular switch that decides whether TDG is involved in DNA repair or in transcriptional activation (Tini *et al.*, 2002). TDG also interacts with some proteins that are not directly involved in transcription. Shimizu *et al.* reported that TDG interacts with XPC-HR23B, a heterodimer that initiates nucleotide excision repair, and that XPC-HR23B promotes release of TDG from abasic sites (Shimizu *et al.*, 2003). Hardeland *et al.* found that SUMO1 and SUMO-2/3 bind to TDG and modify it. They suggested that modification of TDG by SUMO conjugation might be the regulatory mechanism that switches on or off the affinity of TDG for DNA (Hardeland *et al.*, 2002).

*Functional redundancy of Uracil-DNA glycosylases*

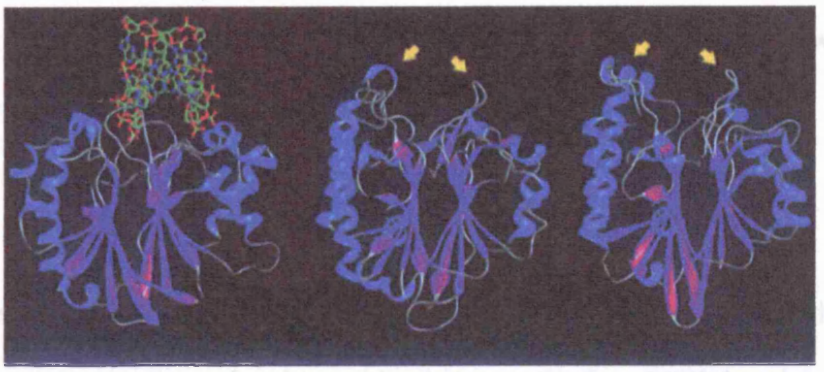
Thymine-DNA glycosylase, MBD4, uracil-DNA glycosylase, and SMUG1 excise uracil from G·U mismatches in mammalian cells. Hence, it raises a question about redundant function among these enzymes. According to the theory of natural selection in evolution, each enzyme should have its own niche. The primary substrate (uracil) of UDG (UNG2) and SMUG1 is identical. However, UDG is localized at the replication site, possibly responsible for misincorporated

uracil and very efficient excision in a single-stranded DNA whereas SMUG1 has a higher efficiency in a double-stranded DNA and may be responsible for repairing uracil derived from deamination of cytosine. Also, it is reported that SMUG1 may have a role in nucleoli formed by loops of DNA containing a cluster of ribosomal RNA genes from several chromosomes, where ribosomal RNAs are transcribed by RNA polymerase I and assembled into the ribosome (Kavli *et al.*, 2002). Also, TDG and MBD4 share a very similar range of substrates, typically thymine mismatched with guanine in CpG sites. MBD4 may have a role in mismatch repair. Furthermore, MBD4 is reported to have 5-methylcytosine DNA glycosylase activity from hemi-methylated CpG sites. However, further study of each enzyme on the function *in vivo* is required to explain the functional redundancy.

## 2.5. Characteristics of Human apurinic/ apyrimidinic endonuclease 1

Apurinic/apyrimidinic (abasic) sites are generated by spontaneous hydrolysis of the glycosydic bond, reaction of DNA glycosylases in base excision repair, and the reaction of damaging agents (e.g. reactive oxygen species, and ionising radiation) (Barzilay and Hickson, 1995). Lindahl and Nyberg estimated the loss of a staggering 2000-10,000 purine bases in each human cell per day (Lindahl and Nyberg, 1972). Human apurinic/apyrimidinic endonuclease 1 (APEX1, also known as HAP1, or Ref-1) has a high affinity for AP sites, and cleaves the phosphodiester bond 5' to the abasic site (Figure 24). APEX1 consists of 318 amino acids (37 kD), and is estimated at ~7 × $10^6$ molecules/cell in HeLa, T lymphoblast and Chinese hamster ovary cells (Chen *et al.*, 1991). It is reported that APEX1 is responsible for more than 95 % of the AP endonuclease activity in mammalian cells (Chen *et al.*, 1991). However, APEX1 has multiple functions in addition to the endonuclease activity, includes 3' phosphodiesterase, 3' phosphatase, 3'-5' DNA exonuclease, RNaseH, and redox activities. Apart from the redox activity, the physiological relevance of these functions is not clear. APEX1 regulates the DNA-binding activity of many transcription factors that are involved in cancer progression, for example, Fos, Jun (Walker *et al.*, 1993; Xanthoudakis et al., 1992), nuclear factor-$_{\kappa}$B (Xanthoudakis *et al.*, 1992), PAX, HIF-1α, HLF, and p53 (Evans *et al.*, 2000) through reduction-oxidation modulation. Therefore, APEX1 links DNA repair and transcription by its two different activities. APEX1 is homologous to *E.coli* exonuclease III, displaying

27 % sequence identity (58 % similarity) (Barzilay and Hickson, 1995). Exonuclease III (Mol *et al.*, 1995) shows a similar structure to DNase I (Lahm and Suck, 1991) despite having less than 20 % sequence homology between them. As expected, the crystal structure of APEX1 showed a similar fold to DNase I and exonuclease III (Gorman *et al.*, 1997) (Figure 23).



**DNase I** **APEX1** **Exo III**

**Figure 23 Comparison of the structure of DNase I, APEX1, and exonuclease III**

**The structure of DNase I is a complex with DNA. Blue indicates α-helices, pink indicates β-strands, and the arrows indicate the loop regions unique to APEX1 and exonuclease III. This figure was taken from Gorman *et al.* (1997).**

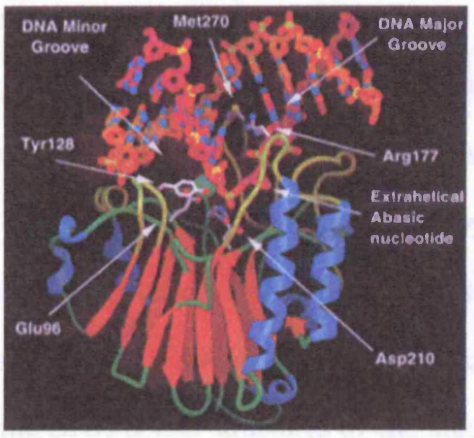



**Figure 24 Crystal structure of human apurinic/apyrimidinic endonuclease 1 complexed with an 11-bp oligonucleotide containing an abasic site**

**The rigid APEX1 with positively charged surface, bends the DNA ~35° and displaces the helical axis ~5Å. APEX1 inserts loops into both the major and minor grooves of DNA, and binds the flipped-out AP site in a binding-pocket. The divalent metal ion (green sphere) is situated at the 5' side of the AP site. This figure was taken from Mol *et al.* (2000).**

The crystal structure of APEX1 bound to the AP site revealed that APEX1 causes kinking of the DNA strands by inserting loops into the major and the minor grooves of DNA, and binds the flipped-out AP site in a separate pocket (Mol *et al.*, 2000) (Figure 24).

From structural studies (Mol *et al.*, 2000), it is known that APEX1 does not exhibit a major conformational transition between the free APEX1 and the DNA-bound APEX1 (~0.7 Å deviation in Cα atom positions between the two forms of APEX1). Therefore, the structure of APEX1 is preformed and does not change on recognition of the abasic site. In contrast, the DNA containing the AP site is severely bent ~35° and kinked ~5 Å relative to the DNA helical axis on binding by APEX1 (Mol *et al.*, 2000). The APEX1 residues Met270 and Met 271 penetrate the DNA minor groove and Arg177 penetrates the major groove to engulf the flipped out abasic site. However, according to a mutational study, none of these residues are responsible for flipping out the abasic site (Mol *et al.*, 2000). This means that the abasic site is possibly flipped out by the gross structural changes induced in the DNA upon enzyme binding. The insertion of the loop into the minor groove is also seen in *Escherichia coli* endonuclease IV (Hosfield *et al.*, 1999). However, the DNA bound structures of endonuclease IV and APEX1 are distinctly different. This suggests that both enzymes may use the loop that inserts into the minor groove to orient themselves relative to the DNA surface and to probe the abasic site at the same time. When the enzymes sense the deformed weak structure with the loop, they may each induce their distinctive distortion in the DNA to specifically bind the flipped out abasic site (Mol *et al.*, 2000).

In base excision repair of G·T mismatches, thymine-DNA glycosylase removes the thymine base to produce an AP site, and, in the presence of APEX1, remains bound to this product DNA. Dissociation of TDG from the abasic site is promoted by APEX1 and is accelerated as the concentration of APEX1 increases (see Section 2.4.3) (Waters *et al.*, 1999). Therefore, APEX1 displaces TDG and then takes over the abasic site and cuts it. The AP endonuclease activity of APEX1 requires divalent cations ($Mg^{2+}$) and it has a $k_{cat}$ of ~500 $min^{-1}$ (Willson III *et al.*, 1995). It has a fast product dissociation rate ($k_{off.product}$ >>10 $s^{-1}$; Strauss *et al.*, 1997), and so the complex of APEX1 and DNA is short lived.

Bennett *et al.* (1997) detected an interaction between APEX1 and DNA polymerase β using yeast two-hybrid analysis. APEX1 leads polymerase β to the DNA damage site via this interaction and allows the polymerase to take over the nicked AP site, and, in short patch repair, removing the 5'-sugar phosphate residue by catalysing a β-elimination reaction (Matsumoto and Kim, 1995). APEX1 also interacts with XRCC1, which in turn interacts with DNA ligase III (Caldecott *et al.*, 1994) and DNA polymerase β (Vidal *et al.*, 2001). This implies that APEX1 is involved in the coordinated mechanism of base excision repair. APEX1 is also involved in long patch base excision repair through its interaction with PCNA and flap endonuclease 1 (Dianova *et al.*, 2001). Therefore, APEX1 may play a switching role from short patch to long patch base excision repair by recruiting flap endonuclease 1 after dissociation of DNA polymerase β (Dianova *et al.*, 2001).

## 2.6. Aims of this project

Human thymine-DNA glycosylase is known to excise thymine mismatched with guanine with a preference for mismatches in CpG sequences. This supports the idea that TDG repairs deaminated 5-methylcytosine because, in humans, methylation of cytosine occurs exclusively in a CpG context, and so subsequent spontaneous deamination generates a G·T mismatch in a CpG site. Saparbaev *et al.* (1998) and Hang *et al.* (1998) reported that TDG excises 3,$N^4$-ethenocytosine; a mutagenic derivative of cytosine that is generated by reaction with vinyl chloride, or with metabolites of lipid peroxidation. Because their results showed that ethenocytosine was apparently removed more efficiently than thymine, these authors claimed that ethenocytosine must be the primary *in vivo* substrate of TDG (Hang *et al.*, 1998; Saparbaev and Laval, 1998). Saparbaev *et al.* (1998) determined the kinetic constants of TDG using Michaelis-Menten kinetics analysis. However, this kinetic model is not appropriate for TDG since TDG is so strongly inhibited by product binding that there is no turnover of the enzyme (Waters *et al.*, 1999) (Scheme 1).

$$TDG + G \cdot T \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} TDG(G \cdot T) \xrightarrow{k_2} TDG(G \cdot AP) \xrightarrow{k_{off}\,(very slow)} TDG + G \cdot AP$$

$$K_d = k_{-1} / k_1$$

**Scheme 1 Kinetic scheme of TDG reaction**

**TDG binds to substrate (G·T mismatched DNA) with rate constant of $k_1$, and dissociates with $k_{-1}$. TDG catalyses cleavage of the glycosydic bond of thymine ($k_2$), and dissociates from the G·AP product site with a very slow $k_{off}$. Dissociation constant is defined as $k_{-1}$ over $k_1$.**

The $k_{cat}$ constant obtained from a Michaelis-Menten experiment measures the overall *turnover* of an enzyme (i.e. the number of molecules of substrate processed per molecule of enzyme per unit time) and in a scheme such as the one shown above would yield little information about the actual catalytic step, $k_2$. Furthermore, because the turnover of TDG is virtually zero, it is actually impossible to carry out a valid Michaelis-Menten analysis of TDG. In addition to this kinetic error, Saparbaev *et al.* used an oligonucleotide that contained a G·T or G·εC mismatch in a GpG context, which is a very poor substrate for thymine excision by TDG (Saparbaev and Laval, 1998). Hang *et al.* (1998) showed that TDG had excision activity towards ethenocytosine, but failed to show the relative activity between thymine and ethenocytosine substrates. The first aim of this project was to measure the dissociation constant ($K_d$) and catalytic constant ($k_2$) of TDG for the removal of thymine and ethenocytosine in various neighbouring sequences using single-turnover kinetics in order to get a clearer understanding of the substrate preference of TDG. Single turnover reactions are the best way to study the kinetics of TDG (and many other glycosylases) since they do not involve the slow product release step.

It is now well established that after removal of the base the subsequent steps in base excision repair are coordinated by a series of interactions where the enzyme of one step recruits the enzyme that caries out the next step (Wilson and Kunkel, 2000). However, the existence of coordination between DNA glycosylases and APEX1 remains controversial. This laboratory has previously shown that APEX1 dissociates TDG from the abasic site and increases turnover of TDG (Waters *et al.*, 1999). There have been suggestions in the literature that the dissociation of TDG is passive and that there is no real coordination between TDG and APEX1

(Hardeland *et al.*, 2001). In this passive mechanism, TDG dissociates on its own, but in the absence of APEX1 quickly re-binds to abasic DNA. When APEX1 is present it binds to free abasic DNA and so prevents TDG from re-binding abasic DNA, thus forcing the glycosylase to bind to G·T DNA and react further. However, the passive dissociation mechanism cannot explain how APEX1 can accelerate the dissociation of TDG greater than the measured rate of dissociation of TDG from abasic DNA, as shown in Figure 17. Therefore, APEX1 must somehow actively displace TDG from abasic DNA, possibly through a direct interaction with TDG. There is little evidence in the literature of protein-protein interactions between TDG and APEX1; only one group detected an interaction (Tini *et al.*, 2002) while other studies looked but failed to find an interaction (Privezentzev *et al.*, 2001). Alternatively, APEX1 may actively displace the TDG by binding to the DNA and distorting it in such a way that weakens the interaction between TDG and the abasic site, thus causing the glycosylase to dissociate. The second aim of this project was to try and deduce how TDG is displaced by the endonuclease. This was done by studying the involvement of the DNA in the displacement reaction and by looking for possible protein-protein interactions between TDG and APEX1.

# 3. MATERIALS and METHODS

# 3.1. Materials

## 3.1.1. Reagents

The chemicals and reagents used for the experiments are listed below.

<Culture medium and Lysis buffer>

| Name | Contents |
|---|---|
| SOC medium | 2 % Bacto-tryptone, 0.5 % of Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose (pH 7.0) |
| Luria-Bertani (LB)-agar plate | 10 mg/ml Bacto-tryptone, 5 mg/ml Bacto-yeast extract, 10 mg/ml NaCl and 15 mg/ml Bacto-agar (pH 7.0) |
| LB medium | 10 mg/ml Bacto-tryptone, 5 mg/ml Bacto-yeast extract, 10 mg/ml NaCl (pH 7.0) |
| Lysis buffer | 50 mM Tris-HCl; pH7.5, 200 mM NaCl, 1 mM ethylenediaminetetraaceticacid (EDTA), 5 % Glycerol and 0.5 % Triton X-100 |

**Table 2 List of medium used for *E.coli* culture and lysis buffer to lysis the cells**

< Fast protein liquid chromatography (FPLC) buffers>

| Name | Contents |
|---|---|
| DNA purification buffer A | 10 mM NaOH and 0.4 M NaCl |
| DNA purification buffer B | 10 mM NaOH and 1.2 M NaCl |
| Mono S buffer A | 25 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulphonicacid (HEPES) (pH 7.6), 1 mM EDTA, 10 % glycerol and 5 mM dithiothreitol (DTT) |
| Mono S buffer B | 25 mM HEPES (pH 7.6), 1 mM EDTA, 10 % glycerol, 5 mM DTT and 1.2 M NaCl |
| Mono Q buffer A | 25 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10 % glycerol and 5mM DTT |
| Mono Q buffer B | 25 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10 % glycerol, 5 mM DTT and 1.2 M NaCl |
| FPLC buffer A | 0.2 M NaCl, 10 mM NaOH, 1 mM EDTA and 0.05 % Triton X-100 |
| FPLC buffer B | 1.2 M NaCl, 10 mM NaOH, 1 mM EDTA and 0.05 % Triton X-100 |
| HiTrap binding buffer | 0.02 M $Na_2HPO_4$ and $NaH_2PO_4$ (pH 7.4) and 0.2 M NaCl |
| HiTrap elution buffer | 0.02M $Na_2HPO_4$ and $NaH_2PO_4$ (pH 7.4), 0.2 M NaCl and 0.5 M imidazole |

**Table 3 List of buffers used for ion exchange chromatography and histidine-trapping column**

<Reaction related buffers>

| Name | Contents |
|---|---|
| Annealing buffer | 25 mM HEPES (pH 7.6), 50 mM KCl and 10 mM $MgCl_2$ |
| Reaction buffer | 25 mM HEPES (pH 7.6), 0.2 mM EDTA, 2 mM DTT, 0.5 mg/ml bovine serum albumin (BSA) and 2.5 mM $MgCl_2$ + either 50 mM or 140 mM KCl |
| Protein buffer | 25 mM HEPES (pH 7.6), 0.5 M KCl, 0.1 mM EDTA, 0.5 mg/ml BSA, 15 % Glycerol and 0.05 % Triton-X-100 |
| Binding buffer | 25 mM HEPES (pH 7.6), 50 mM KCl, 1 mM EDTA, 2 mM DTT, 0.5 mg/ml BSA and 4 % Ficol 400 |
| Quench solution 1 | 0.1 M NaOH and 10 mM EDTA |
| Quench solution 2 | 30 mM Piperidine and 10 mM EDTA |
| Pulldown reaction buffer | 30 mM $NaH_2PO_4$ (pH 7.5), 2 mM $MgCl_2$, 227 mM NaCl, 20 mM imidazole and 27 ng/µl BSA |
| Pulldown washing buffer | 30 mM $NaH_2PO_4$ (pH 7.5), 2 mM $MgCl_2$, 227 mM NaCl and 20 mM imidazole |

**Table 4 List of reaction buffers (kinetic reaction buffer, quenching solution and pulldown reaction buffer) and DNA annealing buffer**

<Gel-electrophoresis related reagents>

| Name | Contents |
|---|---|
| 10 × Tris-borate electrophoresis (TBE) buffer | 0.9 M Tris base (pH 8.0), 0.9 M Boric acid, 20 mM EDTA and 1 % Triton-X-100 |
| 0.5 × TBE buffer | 45 mM Tris base (pH 8.0), 45 mM Boris acid, 1 mM EDTA and 0.05 % Trition-X-100 |
| Tris-acetate electrophoresis (TAE) buffer | 40 mM Tris-acetate and 1 mM EDTA |
| 6% Acrylamide gel | 15 % 40 % Acrylamide : Bis-Acrylamide (19 : 1) and 2.5 % 10 ×TBE buffer |
| 1 % agarose gel | 0.1 μg ethidium bromide/ml TAE buffer and 10 mg agarose /ml TAE buffer |
| 6 × loading buffer | 0.25 % bromophenol blue and xylene cyanol FF and 15 % Ficol 400 |
| MOPS SDS running buffer (Invitrogen) | 10-30 % MOPS, 1-5 % Sodium dodecyl sulfate and 7-13 % Tris |
| MES SDS running buffer (Invitrogen) | 10-30 % Methyl salicylate, 1-5 % Sodium dodecyl sulfate and 7-13 % Tris |

**Table 5 List of gel-electrophoresis related reagents (band shift assay and agarose gel for DNA)**

## 3.1.2. Expression Plasmids

pT7-7 expression plasmids containing the cDNAs of human thymine DNA glycosylase and apurinic/apyrimidinic endonuclease 1 in pT7-7 were kind gifts of J. Jiricny (University of Zürich, Switzerland) (Neddermann *et al.*, 1996) and I. Hickson (Oxford University, UK) (Walker *et al.*, 1993), respectively.

## 3.1.3. Synthesis & Purification of Oligonucleotides

The 34-mer oligonucleotide,

5'-AGCTTGGCTGCAGGCGGACGGATCCCCGGGAATT with the complementary strand containing thymine opposite the underlined guanine was used as the standard G·T glycosylase substrate unless mentioned. The duplexes are described as NpG·X in the following text, where N is the base 5' to the mismatched

(underlined) guanine and X is either thymine, 3,$N^4$-ethenocytosine or uracil opposite the underlined G (see Table 6). These oligonucleotides were used in Section 4.3.

**Table 6 Sequences of 34-mer oligonucleotides**

| Name | X | Sequence |
|---|---|---|
| CpG·T<br>CpG·εC<br>CpG·U | T<br>εC<br>U | 5′-AGCTTGGCTGCAGGCGGACGGATCCCCGGGAATT<br>TCGAACCGACGTCCGXCTGCCTAGGGGCCCTTAA-5′ |
| TpG·T<br>TpG·εC | T<br>εC | 5′-AGCTTGGCTGCAGGTGGACGGATCCCCGGGAATT<br>TCGAACCGACGTCCAXCTGCCTAGGGGCCCTTAA-5′ |
| GpG·εC | εC | 5′-AGCTTGGCTGCAGGGGGACGGATCCCCGGGAATT<br>TCGAACCGACGTCCCXCTGCCTAGGGGCCCTTAA-5′ |
| ApG·εC | εC | 5′-AGCTTGGCTGCAGGAGGACGGATCCCCGGGAATT<br>TCGAACCGACGTCCTXCTGCCTAGGGGCCCTTAA-5′ |

The sequences of oligonucleotide used to study the effect of DNA length on the displacement of TDG by APEX1 are listed in Table 7.

| Name | Sequence |
|---|---|
| Control | 5′-AGCTTGGCTGCAGGCGGACGGATCCCCGGGAATT |
| 6G·T18 | 5′-GCAGGCGGACGGATCCCCGGGAATT |
| 9 G·T18 | 5′-GCTGCAGGCGGACGGATCCCCGGGAATT |
| 15 G·T6 | 5′-AGCTTGGCTGCAGGCGGACGGA |
| 15 G·T5 | 5′-AGCTTGGCTGCAGGCGGACGG |
| 15 G·T4 | 5′-AGCTTGGCTGCAGGCGGACG |
| MINI | 5′-GCTGCAGGCGGACGG |
| 56 Mer | 5′-CGTAGCTGTACATCAGCTTGGCTGCAGGCGGACGGATCCCCGGGAATTACAGATGC |

**Table 7 List of sequences of oligonucleotides used to study effect of DNA in displacement activity of apurinic/apyrimidinic endonuclease 1**

**Only the top strand of each oligonucleotide is shown. The complementary bottom strand contained thymine opposite to the underlined G and was labelled at the 5'-end with a fluorescent molecule (6FAM; see Appendix).**

Oligonucleotides containing ethenocytosine were provided by Dr Tim Waters. Deoxyethenocytidine was synthesized from deoxycytidine following the protocol of Zhang *et al.* (1995), converted to the phosphoramidite and then used to synthesize oligonucleotides using Applied Biosystems 391 DNA synthesizer. The presence of ethenocytosine in the oligonucleotides was confirmed by base composition analysis (Abu and Waters, 2003).

All other oligonucleotides were purchased from Qiagen and Genosys. Since the commercially synthesized oligonucleotides were not pure, all oligonucleotides were purified by ion exchange chromatography at pH 12 (Xu and Swann, 1992) using a Mono Q HR5/5 column (Amersham Biosciences).

Oligonucleotides were dissolved in $H_2O$ at an optical density (260 nm) of approximately 20. The dissolved oligonucleotides were loaded onto a Mono Q column and purified using a gradient of NaCl in 10 mM NaOH run over 20 column volumes. The gradient of the salt concentration varied dependant upon the length and charge (at pH 12) of the oligonucleotide (Xu and Swann, 1992). The oligonucleotide peak was collected and 1 M acetic acid immediately added to adjust the pH to between 5 and 7. The purified DNA sample was desalted using a C18 Sep-Pak cartridge (Waters) eluting with 50 % acetonitrile in water. The fractions containing DNA were dried using a SpeedVac and then re-dissolved in water. The purity of the oligonucleotides was assessed by ion exchange chromatography on a Mono Q column (see Figure 25).

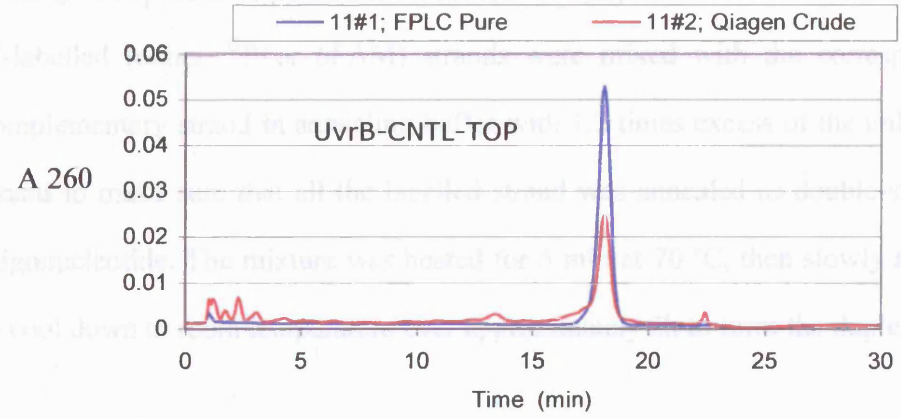



**Figure 25 Typical trace of purification of oligonucleotide using Mono Q column**

**Red line is of the original unpurified oligonucleotide as supplied by Qiagen, the blue line is the oligonucleotide after ion exchange purification.**

The concentration of oligonucleotides was determined by measuring the absorbance at 260 nm and then calculating the concentration using Beer-Lambert's Law (A = ε.c.l) and assuming an extinction coefficient (ε) of $8.2 \times 10^3$ per nucleotide, calculated from Sambrook *et al.* (1989). Oligonucleotide samples were frozen with liquid nitrogen and kept at -20 °C.

### 3.1.3.1. Radioactive labelling of oligonucleotides with [γ-$^{32}$P] ATP

Single-stranded oligonucleotides were radioactively labelled on the 5'-end with [γ-$^{32}$P]-ATP using Ready-To-Go™ T4 Polynucleotide Kinase (Amersham Pharmacia Biotech Inc.).

### 3.1.3.2. Preparation of double-stranded oligonucleotide

5'-labelled (either $^{32}P$ or 6FAM) strands were mixed with the corresponding complementary strand in annealing buffer with 1.2 times excess of the unlabelled strand to make sure that all the labelled strand was annealed as double-stranded oligonucleotide. The mixture was heated for 5 min at 70 °C, then slowly allowed to cool down to room temperature over approximately 2h to form the duplex.

## 3.1.4. Expression and Purification of thymine DNA glycosylase

### *3.1.4.1.* Expression of thymine DNA glycosylase

The pT7.7 expression vector containing the cDNA of thymine DNA glycosylase (pT7-hTDG), was transformed into competent *E.coli* BL21-Gold (DE3) pLysS cells (Stratagene). 50 µl of competent cells were put in a pre-chilled Falcon 2059 tube, and 30 ng of pT7-hTDG DNA was added to the tube and incubated on ice for 20 min. Then, heat-shock was given to the cells at 42 °C for 20 sec in a hot water bath. The tube was immediately placed on ice and left for 2 min. 500 µl of pre-warmed (42 °C) SOC medium was added to the cells. The mixture was shaken horizontally at 37 °C for 1h and then 150 µl of the cells were spread over a pre-warmed (37 °C) LB-agar plate containing 0.1 mg/ml ampicillin and 50 µg/ml of chloramphenicol. The plate was incubated at 37 °C, overnight.

One colony was picked, and then inoculated in 2 ml of LB medium containing 0.1 mg/ml ampicillin and 50 µg/ml of chloramphenicol. The culture was shaken at

37°C for 8h, then 50 µl of the culture was used to inoculate 50 ml of LB medium containing the same antibiotics. The culture was incubated with shaking at 37 °C overnight. Eight 2 l flasks, each containing 400 ml of pre-warmed (37°C) LB medium containing 0.1 mg/ml ampicillin were inoculated with 400 µl of the overnight culture. The flasks were shaken at 37°C until the absorbance of the culture at 595 nm was 0.8. Then, the temperature was lowered to 21°C and isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.6 mM. The flasks were shaken for another 5 hours for expression of TDG and cells were been recovered by centrifugation at 42,000 g for 10 min at 4°C. The final cell pellet (approximately 10 g from 3.2 L of culture) was frozen with liquid nitrogen and stored at –20 °C.

### 3.1.4.2. Lysozyme Lysis of E.coli cells

After thawing the cells, an equal volume of lysis buffer containing 2.5 mg/ml of lysozyme (Sigma) and a cocktail of protease inhibitors (complete EDTA-free, Roche) were added. The mixture was kept on ice for 1h with occasional gentle mixing using a pipette. The lysis mixture was then sonicated for ten cycles of 15 seconds on, 15 seconds off. Then, deoxyribonuclease I (Sigma) and ribonuclease A (Sigma) was added to the lysis mixture to give a final concentration of 20 µg per g of *E.coli* cells. The lysis solution was incubated on ice for a further 1h with occasionally mixing. Finally, the lysis solution was sonicated for ten cycles as before. The mixture was centrifuged at 40,000 g at 4 °C for 30 min and the supernatant containing the soluble TDG was decanted and kept on the ice.

### *3.1.4.3.* Bugbuster Lysis of E.coli cells

The lysozyme method was used for small-scale preparation. However, later cell lysis was done using the following efficient method that gave better extraction of proteins. Five times cell volume of Bugbuster (Novagen) containing 2.5 mM DTT (or 2.5 mM β-mechaptoethanol), 1 mM phenylmethylsulfonylfluoride, a cocktail of protease inhibitors (one tablet of Complete Mini (Roche) per 10 ml lysis mixture) and 25 units/ml of Benzonase (Novagen) was added to the thawed cells. The mixture was incubated for 5 minutes on ice with mixing, then spun down at 16000 g for 20 minutes at 4 °C. Supernatant containing soluble TDG was kept on ice.

### 3.1.4.4. Purification of TDG using a Mono S column

The NaCl concentration of the lysis supernatant was adjusted to 150 mM by dilution with Mono S buffer A and then the solution loaded directly onto a 1ml-Mono S HR5/5 (Amersham Biosciences) anion-exchange column that had been pre-equilibrated with Mono S buffer containing 180 mM NaCl. TDG was eluted using a gradient of 180 mM-360 mM NaCl in Mono S buffer over 20 column volumes. 1 ml fractions of eluate were collected and then analysed by denaturing gel electrophoresis (Section 3.2.1) and by base excision assay on a CpG·T mismatch (Section 3.2.3) to ascertain which contained TDG. Fractions containing TDG (eluted at 320 mM NaCl) were pooled.

### 3.1.4.5. Purification of TDG using a Mono Q column

Pooled TDG fractions after Mono S column purification were further purified through a positively charged 1 ml-Mono Q HR5/5 column (Amersham Biosciences). The NaCl concentration of the pooled Mono S fraction was diluted to 100 mM with Mono Q buffer A before loading onto the Mono Q column. TDG was eluted using a gradient of 60 mM-360 mM NaCl in Mono Q buffer over 20 column volumes. 400 µl fractions were collected and analysed by denaturing gel electrophoresis and by base excision assay on a CpG·T mismatch to determine which contained TDG. Fractions containing pure full-length TDG (eluted at 130 mM NaCl) were pooled.

### 3.1.4.6. Preparation of TDG stock

The NaCl concentration of the pooled TDG fraction was adjusted to 500 mM with 1 M NaCl. The volume of the fraction was then reduced to $\leq$ 250 µl using Centricon-30 (Amicon Inc.) centrifugal filtration device. Finally, 15 % of glycerol and 5 mM DTT were added to the TDG solution. The TDG stock was divided into several tubes, then frozen with liquid nitrogen and stored at –70 °C. To determine the concentration of TDG, five different concentrations of TDG were incubated with $^{32}$P-labelled CpG·T oligonucleotide and then analysed using the band-shift assay (Section 3.2.2) as described in Waters & Swann (1998). The amount of bound and free oligonucleotide was measured by scintillation counting. In these reactions TDG will remove the mismatched thymine but because TDG binds to its product so tightly, all of the TDG in these reactions is bound to DNA. Therefore the amount of TDG is equal to the amount of bound oligonucleotide. A plot of

amount of TDG added versus amount of bound oligonucleotide is linear and so the concentration of TDG can be determined from linear regression of the data (Figure 27).

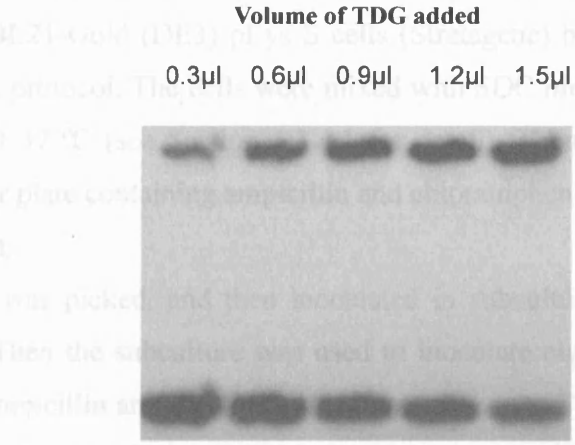



**Figure 26 Typical band-shift assay to determine concentration of TDG using $^{32}$P-ATP rebelled oligonucleotides**

**50 nM oligonucleotide and varied volume of TDG were incubated. Top bands indicate oligonucleotide bound by TDG and bottom bands indicate free oligonucleotides. Each band was cut out to count radioactivity.**

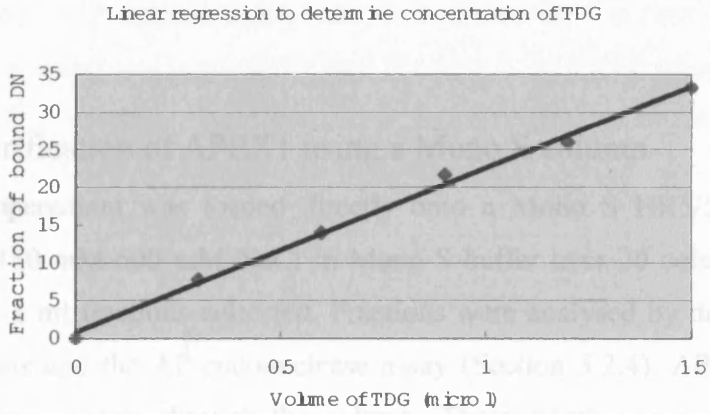



**Figure 27 Linear regression to determine concentration of TDG based on band-shift (Figure 26)**

## 3.1.5. Expression and Purification of Un-Tagged human apurinic/apyrimidinic endonuclease 1 (APEX1)

### 3.1.5.1. Expression of APEX1

The pT7-7 expression vector containing the cDNA of APEX1 was transformed into *E.coli* BL21-Gold (DE3) pLys S cells (Stratagene) by heat-shock following the suppliers protocol. The cells were mixed with SOC medium and incubated on the shaker at 37 ℃ (see Section 3.1.4.1 for detail). Then, the cells were spread over LB-agar plate containing ampicillin and chloramphenicol and incubated at 37 ℃, overnight.

One colony was picked, and then inoculated in subculture containing the same antibiotics. Then the subculture was used to inoculate eight 400 ml-LB medium containing ampicillin and then the culture was shaken at 37°C until absorbance of the 400 ml-culture reached 0.8 at 595 nm. Then, 1 mM of IPTG was added to each flask. The flasks were incubated on the shaker for expression of APEX1 for another 2 h at 37 °C. The cells were recovered and stored as described in Section 3.1.4.1.

The cells were lysed using the lysozyme lysis method described in Section 3.1.4.2.

### 3.1.5.2. Purification of APEX1 using a Mono S column

The lysis supernatant was loaded directly onto a Mono S HR5/5 column. A gradient of 180 mM-600 mM NaCl in Mono S buffer over 20 column volumes was run and 1 ml fractions collected. Fractions were analysed by denaturing gel electrophoresis and the AP-endonuclease assay (Section 3.2.4). APEX1 did not bind and came straight through the column. These fractions were pooled and loaded back on the Mono S column and a gradient of 300 mM-520 mM NaCl was run. Again, fractions were analysed by SDS-PAGE and the AP-endonuclease assay, and the APEX1 containing fractions (eluted at 440 mM NaCl) pooled.

### 3.1.5.3. Purification of APEX1 using a Mono Q column

The APEX1 pool from the second Mono S purification was reduced down to 250 µl using a Vivaspin 4 centrifugal filtration device (Vivascience, Molecular weight cut-off = 10,000 MW). Because APEX1 binds poorly to the Mono Q column, injection of a small volume is important. After reducing the volume the sample was diluted to 40 mM NaCl with Mono Q buffer A, and then loaded onto a Mono Q HR5/5 column. APEX1 was eluted using a salt gradient of 36 mM-216 mM NaCl in Mono Q buffer over 20 column volumes. 400 µl fractions were collected and analysed by SDS-PAGE and the AP-endonuclease assay. Fractions containing pure APEX1 were pooled.

### 3.1.5.4. Preparation of APEX1 stock

The volume of the Mono Q pool was reduced to $\leq$ 250 µl using a Vivaspin 4 centrifugal filtration device (Molecular weight cut-off = 10,000 MW) then 15 % glycerol and 5 mM DTT were added to the sample. To determine the concentration of APEX1, five different concentrations of enzyme were incubated with $^{32}$P-labelled CpG·AP oligonucleotide (prepared by treating CpG·U oligonucleotide with uracil DNA glycosylase as described in Section 3.2.4) and then analysed using the band-shift assay (Section 3.2.2). The amount of bound and free oligonucleotide was measured by scintillation counting and the concentration of APEX1 determined from a plot of amount of APEX1 added versus amount of bound oligonucleotides.

## 3.1.6. Preparation of histidine-tagged human apurinic/apyrimidinic endonuclease 1 pET 100/D-TOPO Expression Vector

The cDNA of APEX1 from the pT7-7 expression vector was subcloned into pET 100/D-TOPO vector (Invitrogen, see Appendix for the map) to attach an N-terminal histidine-tag for the pulldown assay (Section 3.2.6.).

### 3.1.6.1. Amplification of cDNA of APEX1

The primers for PCR were designed so that a 5'-CACC sequence was added in front of the ATG start codon of APEX1. This CACC sequence is required for directional cloning of the PCR fragment into the pET 100/D-TOPO vector *via* the topoisomerase ligation step (see Appendix).

<Primers of APEX1>

Forward; 5′-C ACC ATG CCG AAG CGT GGG AAA AAG GG-3′ (27-mer)

Reverse; 5′-TCA CAG TGC TAG GTA TAG GGT GAT-3′ (24-mer)

(Primers were supplied by MWG.)

```
5'-CACCATGCCGAAGCGTGGGAAAAAGGG-3'  Forward primer
     1 ATGCCGAAGCGTGGGAAAAAGGGAGCGGTGGCGGAAGACGGGGATGAGCT
    51 CAGGACAGAGCCAGAGGCCAAGAAGAGTAAGACGGCCGCAAAGAAAAATG
   101 ACAAAGAGGCAGCAGGAGAGGGCCCAGCCCTGTATGAGGACCCCCCAGAT
   151 CAGAAAACCTCACCCAGTGGCAAACCTGCCACACTCAAGATCTGCTCTTG
   201 GAATGTGGATGGGCTTCGAGCCTGGATTAAGAAGAAAGGATTAGATTGGG
   251 TAAAGGAAGAAGCCCCAGATATACTGTGCCTTCAAGAGACCAAATGTTCA
   301 GAGAACAAACTACCAGCTGAACTTCAGGAGCTGCCTGGACTCTCTCATCA
   351 ATACTGGTCAGCTCCTTCGGACAAGGAAGGGTACAGTGGCGTGGGCCTGC
   401 TTTCCCGCCAGTGCCCACTCAAAGTTTCTTACGGCATAGGCGAGGAGGAG
   451 CATGATCAGGAAGGCCGGGTGATTGTGGCTGAATTTGACTCGTTTGTGCT
   501 GGTAACAGCATATGTACCTAATGCAGGCCGAGGTCTGGTACGACTGGAGT
   551 ACCGGCAGCGCTGGGATGAAGCCTTTCGCAAGTTCCTGAAGGGCCTGGCT
   601 TCCCGAAAGCCCCTTGTGCTGTGTGGAGACCTCAATGTGGCACATGAAGA
   651 AATTGACCTTCGCAACCCCAAGGGGAACAAAAAGAATGCTGGCTTCACGC
   701 CACAAGAGCGCCAAGGCTTCGGGGAATTACTGCAGGCTGTGCCACTGGCT
   751 GACAGCTTTAGGCACCTCTACCCCAACACACCCTATGCCTACACCTTTTG
   801 GACTTATATGATGAATGCTCGATCCAAGAATGTTGGTTGGCGCCTTGATT
   851 ACTTTTTGTTGTCCCACTCTCTGTTACCTGCATTGTGTGACAGCAAGATC
   901 CGTTCCAAGGCCCTCGGCAGTGATCACTGTCCTATCACCCTATACCTAGC
                     Reverse Primer  3'-TGCTGAGTGATAGTGGGATATGGATCG
   951 ACTGTGA
       TGATGT-5'
```

**Figure 28 cDNA sequence of human apurinic/apyrimidinic endonuclease 1 showing where PCR primers anneal**

62.5 ng of pT7-7-APEX1 vector, 0.2 μM of each primer and 125 μM of each dNTPs were added to 50 μl of PCR mixture, following the manufacturers protocol. Lastly, Vent DNA polymerase was added to the mixture. PCR was carried out using 25 cycles of 94 °C for 30 sec, 55 °C for 45 sec, and 75 °C for 1 min. At the end of the last cycle the temperature was held at 75 °C for 10 min and then decreased to 4 °C.

The PCR product was examined on a 1 % agarose gel run in TAE buffer at 20 V/cm for 40 min and then stained with ethidium bromide.

The major PCR band, which was the size expected for the cDNA of APEX1 (961 bp), was cut out and extracted from the gel using a QIAquick gel extraction kit (Qiagen).

### 3.1.6.2. Ligation of APEX1 into pET100/D-TOPO vector and Transformation into One Shot® TOPO10 cells

8.5 ng of the purified PCR product was mixed with 10 ng of pET100/D-TOPO® vector (Invitrogen) following the manufacturers instructions, which recommend a 5-fold molar excess of PCR product over vector. The mixture was mixed gently and incubated for 5 min at room temperature. Then, the reaction was put on ice.

3 μl of the ligation mixture was put into a vial containing 50 μl of One Shot® TOPO10 cells (Invitrogen) and mixed gently. The mixture was kept on ice for 15 min and then placed in a hot bath (42 °C) for 30 sec. The vial was immediately put back on ice for 2 min. 250 μl of room temperature SOC medium was added and the transformation mixture shaken at 37 °C for 30 minutes. 150 μl of the culture was spread on a pre-warmed LB agar plate containing 0.1 mg/ml ampicillin and incubated overnight at 37 °C. The next day, one colony was picked to inoculate 5 ml of LB medium containing 0.1 mg/ml ampicillin, and shaken overnight at 37 °C.

### 3.1.6.3. Verification of pET-100/D-TOPO-His-APEX1 Vector

The 3 ml overnight culture was used for a mini-scale preparation of plasmid DNA using a QIAprep spin miniprep kit (Qiagen) following the manufacturers instructions.

The purified plasmid DNA was digested with a restriction enzyme, SacI (New England Biolabs), to confirm the presence of the insert and that it was in the correct orientation (see Section 4.2.).

The clone that contained the insert in the right direction, was further cultured in 100 ml of LB medium containing 0.1 mg/ml ampicillin to prepare more plasmid DNA for sequencing and storage. After overnight culture at 37 °C, cells were harvested by centrifugation and then plasmid DNA purified using a HiSpeed Plasmid Maxi Kit (Qiagen) following the manufacturers protocol. DNA sequencing of the purified plasmid (University of Cambridge sequencing service) confirmed that the insert was successfully ligated in pET100/D-TOPO.

## 3.1.7. Expression and Purification of Histidine-Tagged human apurinic/apyrimidinic endonuclease 1 (His-APEX1)

### 3.1.7.1. Expression of His-APEX1

The His-APEX1 pET expression vector was transformed into BL21-Gold (DE3) pLys S cells and His-APEX1 expressed as described in Section 3.1.5.1.

### 3.1.7.2. Purification of histidine-tagged APEX1

The cells were lysed with lysozyme as described in Section 3.1.4.2. except that the lysis buffer did not contain EDTA. The lysis supernatant was directly loaded onto a nickel-charged 1 ml HiTrap™ Chelating HP column (Amersham Bioscience) that had been equilibrated with 30 mM imidazole in HiTrap buffer. The histidine-tagged APEX1 was eluted using a gradient of 30 mM-160 mM imidazole in HiTrap buffer over 10 column volumes. 1 ml fractions were collected and then analysed for histidine-tagged APEX1 by SDS-PAGE and AP-endonuclease activity. Fractions containing His-APEX1, which eluted at 70 mM imidazole, were pooled and kept at 4 °C.

The pool containing His-APEX1 was diluted to 156 mM NaCl with Mono S buffer A and then loaded onto a 1 ml-Mono S column. The histidine-tagged APEX1 was eluted with a gradient of 160 mM-460 mM NaCl in Mono S buffer over 20 column volumes, collecting 1 ml fractions. Fractions were analysed by SDS-PAGE and the AP-endonuclease assay. The His-APEX1 containing fractions were pooled (eluted at 290 mM NaCl), and the volume reduced with a Vivaspin 4 centrifugal filtration device (molecular weight cut-off = 10,000 MW). The concentration of His-APEX1 was estimated using the Bio-Rad protein dye-binding assay (Bradford, 1976).

# 3.2. Experimental Procedures

## *3.2.1.* Denaturing Gel Electrophoresis (SDS-PAGE) of Proteins

10 % acrylamide Bis-Tris, NuPAGE® gels (Invitrogen) were used to separate proteins. Protein samples were prepared in NuPAGE® LDS Sample Buffer (Invitrogen) that contained 5 % β-mercaptoethanol, then heated for 5min at 90 °C. Gels were run with denaturing, NuPAGE® 2-(*N*-morpholino) ethane sulfonic acid SDS Running Buffer (Invitrogen) at 30 V/cm gel. Gels were stained using Silver Stain Plus Kit (Biorad) following the manufacturers instructions.

## 3.2.2. Band-shift assay

1.5 mm-thick nondenaturing 6 % polyacrylamide (ratio of Acrylamide to Bis-acrylamide = 19:1) gels were pre-run at 20 V/cm in 0.5 × TBE for 1h to elute ions and un-polymerised acrylamide. TDG was incubated with the appropriate $^{32}$P-labelled DNA in binding buffer (see Table 4) at room temperature for 2 hours. 10 µl of each binding mixture was loaded onto a pre-run gel and electrophoresis performed at 11 V/cm in 0.5 × TBE buffer for 1 h at a temperature of between 15 and 20°C. Following electrophoresis, the gel was dried and exposed to an X-ray film (Kodak).

## *3.2.3.* Assay of Glycosylase Activity

NpG·X (see Table 6 for explanation of nomenclature) oligonucleotides, 5'-labelled in the X containing strand with either $^{32}P$ or 6FAM, were mixed in reaction buffer with TDG. Reactions were carried out at room temperature. The excision reaction was terminated by adding to quench solution 1 (quench solution 2 for ethenocytosine oligonucleotides since the glycosidic bond of ethenocytosine is susceptible to hydrolysis in strong basic condition) and then heating at 90 °C for 30 min to break phosphodiester bond at the abasic site. Then, the cleaved DNA was separated from full length unreacted oligonucleotide by cation-exchange chromatography using a 0.1 ml POROS HQ column (PerSeptive Biosystems). Typically, the DNA was eluted by a gradient elution of 420 mM-600 mM NaCl in FPLC buffer A and B over 32 column volumes but sometimes (*e.g.* when using oligonucleotides of a different length) the gradient was adjusted to ensure that product and substrate peaks were adequately separated (see Figure 29).

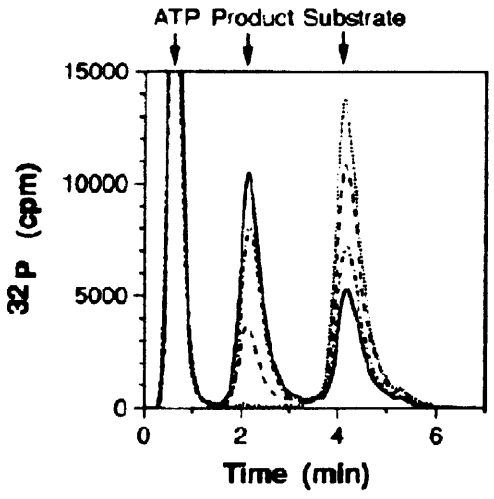



**Figure 29 Typical trace of separation of oligonucleotide substrate and product from a glycosylase reaction by ion exchange chromatography and radioactive monitoring**

**The first peak corresponds to un-incorporated free $^{32}P$-ATP (from the labeling reaction), the second corresponds to reacted short oligonucleotides (product) and the third intact oligonucleotides. The figure was taken from Waters and Swann (1998).**

Radiolabelled DNA was detected by Cerenkov counting using a Berthold LB 506 C-1 monitor and fluorescently labelled DNA was detected by fluorescence emission at 525 nm (excitation at 494 nm) using a RF-10A $_{XL}$ detector (Shimazu). The peaks were quantified by integration.

### ~~3.2.3.1.~~ Measurement of $K_d$ and $K_2$ Kinetic Constants for TDG Base Excision

As discussed in the introduction, the TDG base excision reaction can be described by the scheme shown below (where E stands for enzyme, S for substrate and P for product).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} EP \xrightarrow{k_{off}\,(veryslow)} E + P$$

With most TDG substrates, $k_{off}$ is so slow (~ $10^{-5}$ $s^{-1}$) that it can be assumed to be essentially zero and the product release step ignored. The rates of reaction for each of the components in the scheme can thus be written as follows.

$d[E]/dt = -k_1*[E]*[S] + k_{-1}*[ES]$
$d[S]/dt = -k_1*[E]*[S] + k_{-1}*[ES]$
$d[ES]/dt = k_1*[E]*[S] - k_{-1}*[ES] - k_2*[ES]$
$d[EP]/dt = k_2*[ES]$

The $k_2$ constant gives the rate of the actual base excision step and can be determined from the maximum rate of reaction achieved when the enzyme is saturated with substrate. The ratio of $k_{-1}/k_1$ gives the binding dissociation constant ($K_d$) for substrate binding to the enzyme and can be determined from the dependence of reaction rate on enzyme and substrate concentration. To measure values of $k_{-1}/k_1$ and $k_2$ single turnover reaction rates were measured for equimolar amounts of TDG and oligonucleotide at seven different TDG/substrate concentrations. Reactions were carried out as described in the above section except that the reaction buffers contained 140 mM rather than 50 mM KCl. Reaction rates were measured by removing samples at different times from the same reaction mixture, quenching and then determining the extent of base excision as described above. Data for all seven concentrations were fitted simultaneously to the rate equations shown above using the differential equation

solving program Berkeley Madonna (version 8.0.1; www.berkeleymadonna.com). The program was used to obtain the best fit of the theoretical lines to all of the experimental data by varying the values of $k_1$, $k_{-1}$ and $k_2$. Approximate values of $K_d$ and k2 from preliminary experiments were used to adjust concentrations so that they were 0.1, 0.2, 0.5, 1, 2, 5 and 10 times the approximate value of $K_d$ and to adjust the time that samples were removed so that five time points were obtained for the first ~30% of the reaction was measured.

### 3.2.3.2. Inhibition Assay

In order to study the affinity of TDG for thymine or 3,$N^4$-ethenocytosine bases, the reaction of CpG·T was measured in the presence of thymine or ethenocytosine base to see if either base could inhibit excision. Since up to 5 mM of both bases had no effect, single stranded 34-ner oligonucleotide containing either thymine (bottom strand of CpG·T) or ethenocytosine (bottom strand of CpG·eC) were tested as inhibitors. Because TDG does not react with single-stranded DNA (Abu and Waters, 2003), they were expected to act reversible inhibitors. 25 nM of fluorescent labelled CpG·T mismatch oligonucleotide, and either no inhibitor, 250 nM of thymine-containing single stranded oligonucleotide or 60 nM of 3,$N^4$-ethenocytosine-containing single stranded oligonucleotide were mixed with 25 nM TDG in reaction buffer containing 140 mM KCl and incubated at room temperature. Samples were removed at various times, quenched and analysed by ion exchange chromatography as described in Section 3.2.3.

## 3.2.4. Assay of AP Endonuclease Activity

500 nM of 34-mer oligonucleotide containing a U·G mismatch and 5 nM of *Herpes simplex* uracil DNA glycosylase (a kind gift from Renos Savva, Birkbeck College, UK) were pre-incubated in reaction mixture at room temperature for 1 h to allow uracil DNA glycosylase to excise all of the uracil to produce the abasic site substrate of APEX1. Samples of diluted fractions from the APEX1 purification were added to the abasic DNA, incubated for 2 min at room temperature and then the mixture was loaded directly onto a 0.24 ml Mini-Q (Amersham Biosciences cation-exchange chromatography column. Cleaved abasic oligonucleotide was separated from intact, full-length abasic oligonucleotide using a gradient of 500 mM-650 mM NaCl in FPLC buffer over 13 column volumes. The absorbance of substrate and product was monitored at 260 nm.

## 3.2.5. Assay of TDG Displacement by APEX1

In this assay, the reaction of TDG with an excess of substrate can be enhanced by addition of APEX1, which displaces the TDG from its abasic product. Typically, 10 nM of TDG and 100 nM (15 nM fluorescent labelled and 85 nM non-labelled) of oligonucleotide substrate and variable amounts of APEX1 were incubated at room temperature in reaction buffer. Samples were removed at various times and put in quench solution 1 (quench solution 2 for ethenocytosine oligonucleotides) followed by heating at 90 °C for 30 min to break the phosphodiester bond at abasic sites. Cleaved DNA therefore includes both abasic oligonucleotides cut by APEX1 and abasic oligonucleotides cleaved by this hot hydroxide treatment. Cleaved oligonucleotides were separated from intact substrate by ion exchange on a POROS HQ column as described in Section 3.2.3.

### 3.2.5.1. Displacement Assay with different length of oligonucleotides

When APEX1 is present, TDG is dissociated from the abasic site and allowed to catalyse more substrates. However, we do not know how APEX1 approaches the complex of TDG-abasic site nor how the DNA strands are involved during the displacement. To study the influence of flanking DNA length on APEX1 displacement of TDG from abasic sites, eight oligonucleotides of varied length and varied position of the mismatch were prepared (see Table 7). The sequences of oligonucleotides were designed using the DNase I footprint experiment of TDG carried out by Schärer, *et al.* as a guide (Schärer *et al.*, 1997).

20 nM TDG was incubated with 24 nM of fluorescent labelled (6-FAM) oligonucleotide in reaction buffer containing 50 mM KCl for 1h at room temperature. During this pre-incubation TDG removes the mismatched thymine from the FAM labelled G·T substrate so that all of the TDG is now bound to the abasic product of that substrate. A sample of this was analysed by HPLC (monitoring fluorescence) to confirm that the TDG reaction on the 6FAM oligonucleotides completed. Then, 20 nM of $^{32}P$ labelled 34-mer CpG·T 'reporter' oligonucleotide, followed by none or 50 nM of APEX1, were added to the reaction mixture to study how much TDG was dissociated from the FAM labelled substrate to process the $^{32}P$ labelled reporter oligonucleotide. Samples were removed at various time intervals, quenched and then the amount of thymine removed from the reporter oligonucleotide analysed as described in Section 3.2.3.

## 3.2.6. Pulldown Assay

To prepare the nickel-charged resin, Ni-NTA agarose (Qiagen) was washed with water and spun down to remove supernatant. Then, the resin was re-charged with nickel by mixing with 0.1 M Nickel sulphate, washed with water and then incubated in pulldown reaction buffer for 15 min at 4 °C. 30 µl of pulldown reaction buffer containing either 500 nM TDG, 1.75 µM histidine-tagged APEX1, or both, was incubated for 15 min at 4 °C. 9 µl samples of each mixture were taken out and mixed with 3 µl of 4 × NuPAGE® LDS Sample Buffer for SDS-PAGE analysis. 2 µl of prepared agarose resin was added to the remainder of each reaction mixture and then incubated for 1h 30min at 4 °C. After the incubation, the mixtures were spun down and 9 µl of supernatant was added to 3 µl of 4 × NuPAGE® LDS Sample Buffer. The resin was then washed twice with 45 µl of washing buffer. After discarding supernatant, 10 µl of 1 × NuPAGE® LDS Sample Buffer was added to the pellet. All samples were analysed by SDS-PAGE as described in Section 3.2.1.

## 3.2.7. Assay for Protein-Protein Interaction using Isothermal Titration Calorimetry (ITC; MicroCal™)

Isothermal titration calorimetry (VP-ITC) was used to detect interaction of TDG and His-APEX1. 2 ml of 6 µM TDG and 1 ml of 60 µM His-APEX1 were prepared for one experiment. After purification of the proteins, they were extensively dialysed in the same Mono Q buffer containing 200 mM NaCl. TDG was put in the sample cell and the temperature equilibrated to 30 °C. Then 15 µl of His-APEX1 was injected every 2.5 min, repeating for 17 times. The heat produced after every injection was monitored.

# 4. RESULTS and DISCUSSION (1); Kinetic Study of Human Thymine-DNA Glycosylase

- ✧ Purification of Thymine-DNA Glycosylase
- ✧ Purification of Apyrimidinic/apurinic Endonucleas 1
- ✧ Kinetic Study of Human Thymine DNA Glycosylase

# 4.1. Purification of Thymine-DNA Glycosylase

The cDNA of human thymine-DNA glycosylase was expressed in *E.coli* BL21-Gold (DE3) pLysS cells using pT7-7 vector (see Figure 30). The transformed cells were allowed to grow in the presence of ampicillin. Expression of the desired protein was induced by IPTG and at lower temperature (21°C). Approximately 8 g of cells were harvested from 3.2 l of culture.

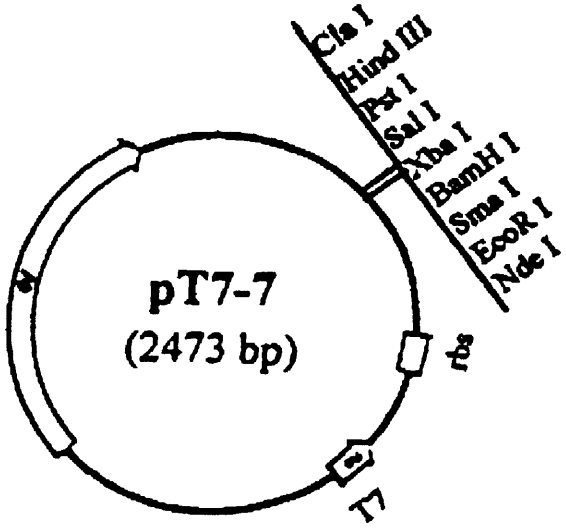



**Figure 30 The map of plasmid, pT7-7**

**It contained cDNA of thymine DNA glycosylase or human apurinic/apyrimidinic endonuclease 1 between EcoR I and Hind III restriction sites.**

The cells were lysed with lysozyme and sonication, and then the mixture was centrifuged (see Materials and Methods).

The supernatant from the cell lysis was directly loaded onto a Mono S column (see Section 3.1.4.4). Initially, purification of TDG suffered from substantial degradation as shown in Figure 31.

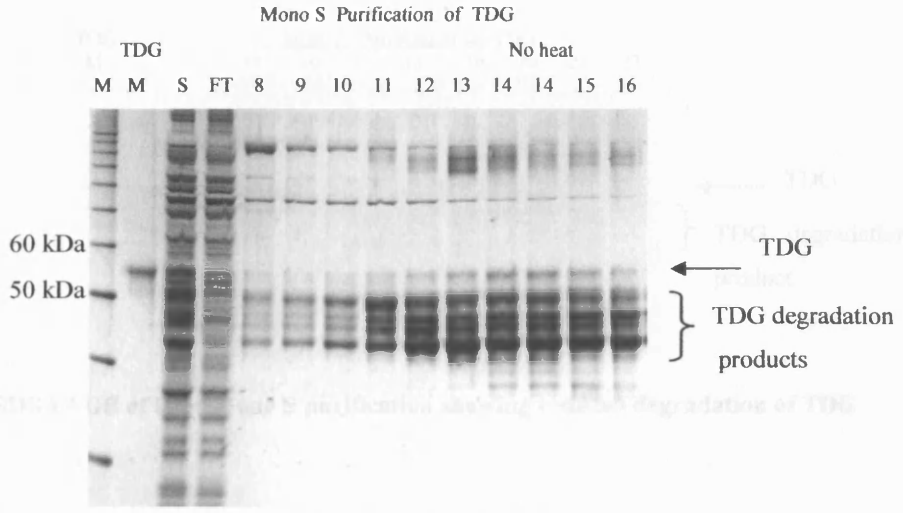



**Figure 31 SDS-PAGE showing purification of TDG using Mono S column**

**M stands for marker, S for supernatant of lysis mixture, FT for flow-through and numbers are collected 1 ml fractions. One sample of Fraction 14 was not heated before loading on the gel to examine if heating caused the degradation of TDG.**

The trace of TDG purification with Mono S showed a broad peak (Fraction 11 – 16 in Figure 31) using a gradient of 180 mM-360 mM NaCl. As indicated in the graph, majority of TDG was degraded and a small fraction of full-length TDG (eluted at 320 mM NaCl) was present. Heating samples (at 90 ℃ for 5 min) before loading onto the SDS-PAGE was not the major cause of the degradation (see Fraction 14). Later, purification methods were improved by adding extra protease inhibitors and by using the faster Bugbuster lysis method (Section 3.1.4.3).

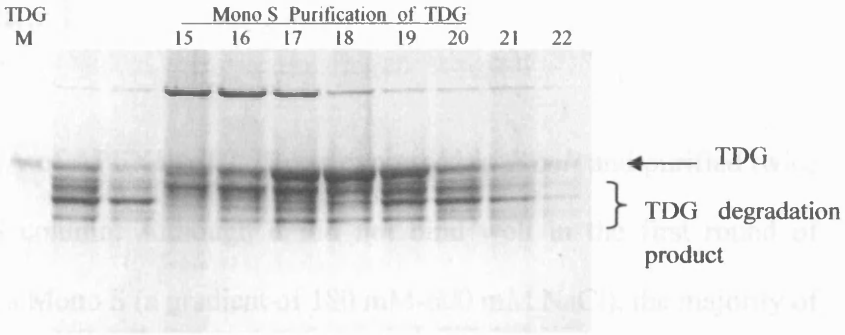



**Figure 32 SDS-PAGE of later Mono S purification showing reduced degradation of TDG**

**Mono S fractions containing intact TDG were pooled and then loaded onto a Mono Q column for further purification (see Figure 33).**

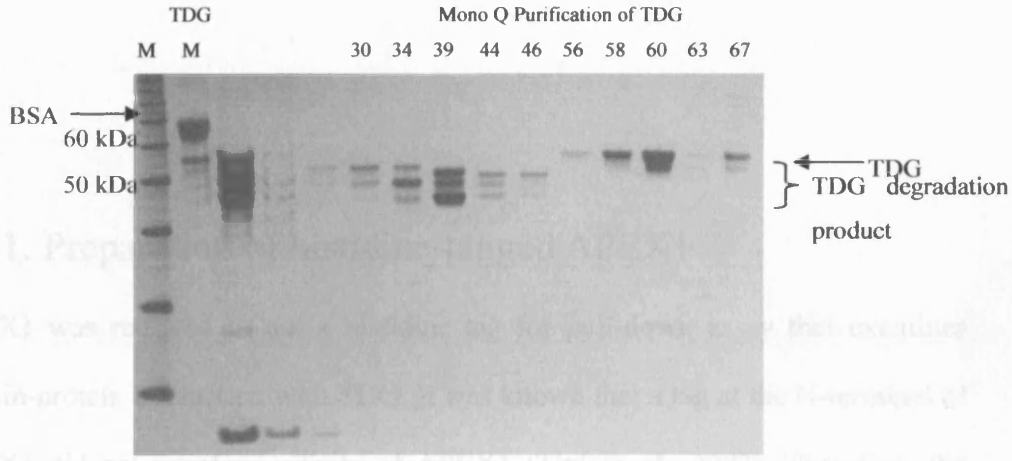



**Figure 33 SDS-PAGE showing purification of TDG using Mono Q column**

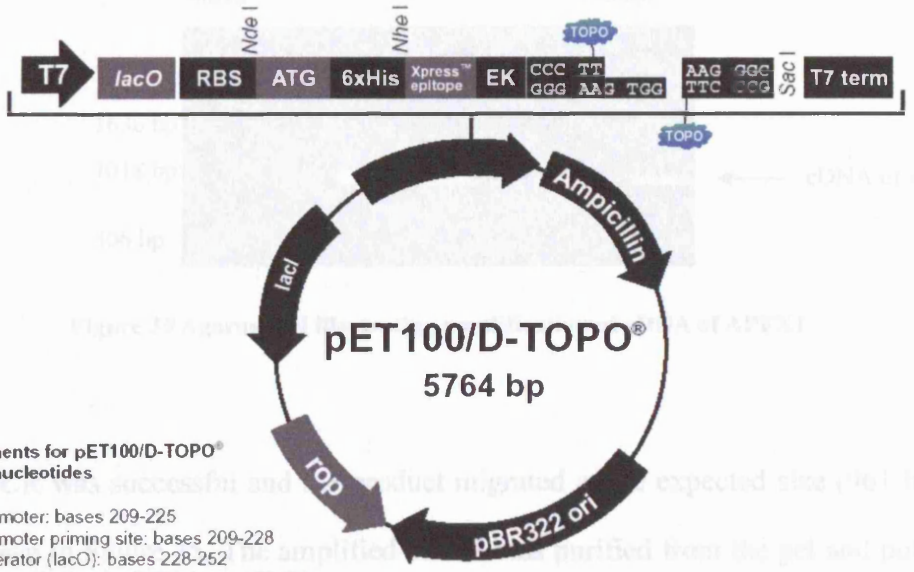
**TDG was eluted from the Mono Q column at 130 mM NaCl. As seen in Figure 33, full-length TDG was well separated from the degradation products. Fractions were examined by TDG assay and fractions containing full-length TDG were pooled. This final step of purification yielded pure TDG. The volume was reduced to give approximately 25 μM of TDG.**

# 4.2. Purification of Apyrimidinic/apurinic Endonucleas 1

Non-tagged cDNA of APEX1 (pT7-7) was expressed in *E.coli* and purified twice using a Mono S column. Although it did not bind well in the first round of purification with a Mono S (a gradient of 180 mM-600 mM NaCl), the majority of contaminants were eliminated. In the second Mono S purification (a gradient of 300 mM-520 mM NaCl), APEX1 was eluted at 440 mM NaCl. Finally, pure APEX1 was yielded from purification using a Mono Q column (a gradient of 36 mM-216 mM NaCl), which eliminated contaminants that bound to the column (see Section 3.1.5.2 and 3.1.5.3).

## 4.2.1. Preparation of histidine-tagged APEX1

APEX1 was required to put a histidine tag for pull-down assay that examines protein-protein interaction with TDG. It was known that a tag at the N-terminal of APEX1 did not interfere activity of APEX1 (Tini *et al.*, 2002). Therefore, the cDNA of APEX1 (pT7-7) was subcloned into pET100/D-TOPO (see Figure 34).




**Comments for pET100/D-TOPO®**
**5764 nucleotides**

T7 promoter: bases 209-225
T7 promoter priming site: bases 209-228
lac operator (lacO): bases 228-252
Ribosome binding site (RBS): bases 282-288
Initiation ATG: bases 297-299
Polyhistidine (6xHis) region: bases 309-326
Xpress™ epitope: bases 366-389
EK recognition site: bases 375-389
TOPO® recognition site 1: bases 396-400
Overhang: bases 401-404
TOPO® recognition site 2: bases 405-409
T7 reverse priming site: bases 466-485
T7 transcription termination region: bases 427-555
*bla* promoter: bases 856-954
Ampicillin (*bla*) resistance gene: bases 955-1815
pBR322 origin: bases 2022-2757
*ROP* ORF: bases 3001-3192 (complementary strand)
*lacI* ORF: bases 4507-5595 (complementary strand)

**Figure 34 The map of pET100/D-TOPO**

**The cDNA of human apurinic/apyrimidinic endonuclease 1 was inserted in this plasmid in order to add the histidine tag at the N-terminal.**

The directional TOPO cloning system (Invitrogen) was employed (see Appendices). First of all, the cDNA of APEX1 in pT7-7 was amplified using primers that added 5'-CACC-3' in front of a start codon, ATG, in PCR (see Section 3.1.6.1). The PCR reaction was examined in a 1 % agarose gel (Table 5).

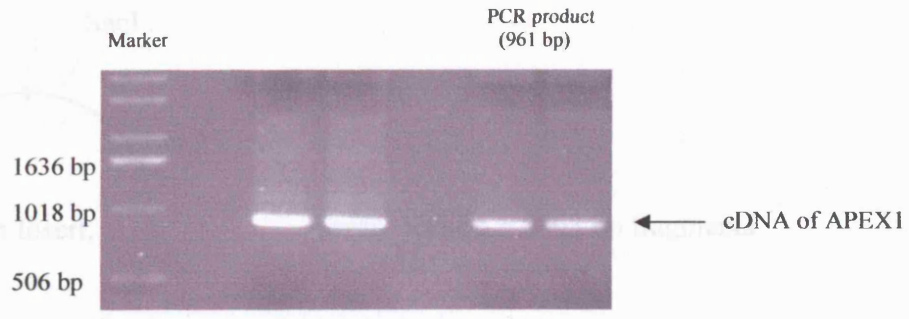



**Figure 35 Agarose gel illustrating amplification of cDNA of APEX1**

The PCR was successful and the product migrated at the expected size (961 bp), as shown in Figure 35. The amplified cDNA was purified from the gel and put in the ligation mixture with pET 100/D-TOPO (Section 3.1.6.2). After transformation, a few clones were cultured and plasmid DNAs were prepared to verify the ligation by restriction digestion using SacI (Section 3.1.6.3).

If the ligation was successful SacI would cut at two sites in the construct; one is 54 base pairs into the APEX1 insert and the second is 15 base pairs downstream of the insert (see Figure 36 below). Thus, products of digestion were expected to be 922 bp and 5803 bp. After incubation for 2h at 37 °C the SacI digestion was loaded onto a 1 % agarose TAE gel and the gel run at 20 V/cm for 45 min. Ethidium bromide staining of the gel revealed two strong bands at ~6000 bp and ~900 bp, indicating that the plasmid contained the APEX1 cDNA inserted in the correct direction.

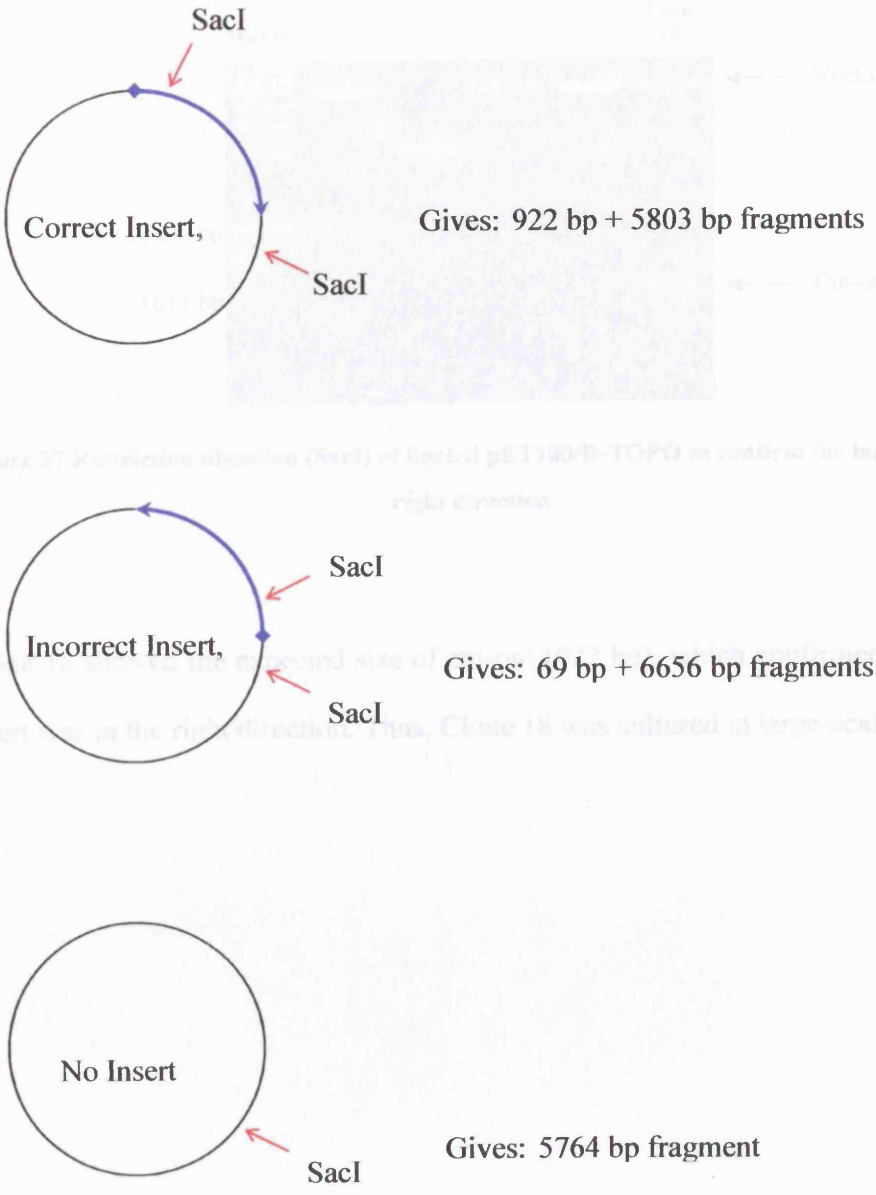



**Figure 36 Expected sites of restriction digestion of pET-100/D-TOPO-His-APEX1 Vector with SacI**

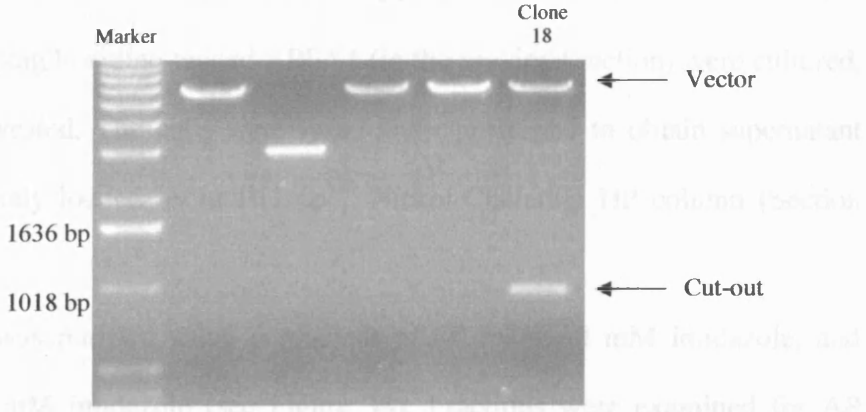



**Figure 37 Restriction digestion (SacI) of ligated pET100/D-TOPO to confirm the insert in the right direction**

Clone 18 showed the expected size of cut-out (922 bp), which confirmed that the insert was in the right direction. Thus, Clone 18 was cultured in large-scale.

## 4.2.2. Purification of histidine-tagged APEX1 (His-APEX1)

*E.coli* expressing histidine-tagged APEX1 (in the previous section) were cultured, and then harvested. The cells were lysed and centrifuged to obtain supernatant that was directly loaded on to HiTrap™ Nickel-Chelating HP column (Section 3.1.7.2).

His-APEX1 was purified using a gradient of 30 mM-160 mM imidazole, and eluted at 70 mM imidazole (see Figure 38). Fractions were examined for AP endonuclease activity and fractions containing histidine-tagged APEX1 were pooled.

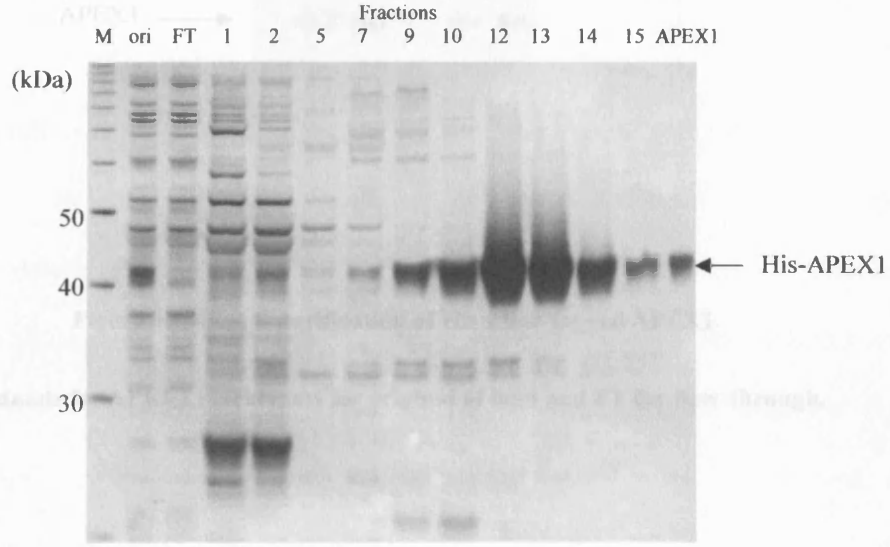



**Figure 38 SDS-PAGE showing Purification of histidine-tagged APEX1**

**A; supernatant of lysis was directly loaded onto the histidine-trapping HiTrap column. Fraction 10-15 was pooled. M stands for marker, ori for origin of load, and FT for flow-through.**

Although this step gave almost pure APEX1, the pooled fractions were loaded onto Mono S column for further purification (Section 3.1.7.2). Fractions from the Mono S column were analysed by SDS-PAGE (see Figure 39) and assayed for

AP-endonuclease activity. Some His-APEX1 went straight through the column, however, the His-APEX1 (which eluted at 290 mM NaCl) had no visible contaminating protein bands. Fractions containing pure His-APEX1 were pooled, and then volume of the pool was reduced to give approximately 25 or 100 μM of His-APEX1.

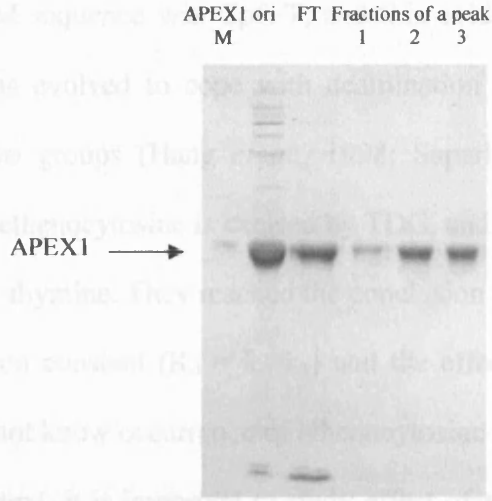



**Figure 39 Mono S purification of Histidine-tagged APEX1**

**M stands for APEX1 marker, ori for original of load and FT for flow-through.**

# 4.3. Kinetic Study of Human Thymine DNA Glycosylase

TDG has a relatively broad range of substrates and a sequence preference. The sequence preference is in the order of CpG·T >> TpG·T > GpG·T > ApG·T (Cross and Bird, 1995; Sibghat-Ullah *et al.*, 1996; Waters and Swann, 1998). The most preferred sequence was CpG·T, and this evidence supports the hypothesis that TDG was evolved to cope with deamination of 5-methylcytosine in CpG sites. The two groups (Hang *et al.*, 1998; Saparbaev and Laval, 1998) have reported that ethenocytosine is excised by TDG, and it would be a better substrate for TDG than thymine. They reached the conclusion with $k_{cat}$ and $K_m$, but not with the dissociation constant ($K_d = k_{-1}/k_1$) and the effect of neighbouring sequence. Since we do not know occurrence of ethenocytosine has a bias toward a particular sequence context, it is important to study effect of adjacent sequence in cleavage of ethenocytosine. In this project, $K_d$ and $k_2$ (see Scheme 2) for thymine and ethenocytosine mismatched with guanine in various sequence contexts (i.e. CpG·T, TpG·T, CpG·εC, TpG·εC, GpG·εC and ApG·εC) were measured, to characterize the excision reaction of TDG.

Kinetics is expressed in terms of the rates of chemical reactions based on the Law of Mass Action. The law, proposed by Guldberg and Waage in 1867, says that the rate of reaction is proportional to the product of the concentration of each reactant (Palmer, 1995). A degree of how fast the reaction takes place is expressed as the rate constant, k. Thymine DNA glycosylase shows a single turnover reaction, which means that one molecule of TDG reacts with one molecule of substrate, and

TDG stays bound with its product, an abasic site. Therefore, the schematic mechanism of the reaction is:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} EP$$

**Scheme 2 The kinetic mechanism of enzymatic reaction**

**At the first step, the enzyme (E) and the substrate (S) form a complex (ES). This step is considered to be fast and reversible. The rate constant of formation of the ES is depicted as $k_1$, and of disassociation of the ES indicated as $k_{-1}$. The dissociation constant, $K_d$ is expressed by $k_{-1}/k_1$. In the second step, the enzyme catalyses a chemical reaction with the catalytic constant, $k_2$ forming a complex of enzyme-product (EP). This scheme was derived from the theory developed by Brown in 1902, L. Michaelis and M. L. Menten in 1913.**

## 4.3.1. Comparison of a conventional Lineweaver-Burk plot with Berkeley Madonna for determination of $K_M$ and $k_{cat}$

In order to compare two different methods (a conventional Lineweaver-Burk plot and Berkeley Madonna differential equation solving program) for calculating $k_{cat}$ and $K_M$, a fictional set of data for an enzyme reaction that follows Michaelis-Menten theory (Scheme 3) was calculated (presented in Table 8).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

**Scheme 3 The kinetic mechanism of enzymatic reaction based on the theory developed by Brown in 1902, then L. Michaelis and M. L. Menten in 1913**

**Brown in 1902, then L. Michaelis and M. L. Menten in 1913**

**5 nM Substrate**

| Time (sec) | Fraction of Substrate reacted | Concentration of Substrate reacted (nM) |
|---|---|---|
| 0.0 | 0.0000 | 0.000 |
| 11.7 | 0.0250 | 0.125 |
| 23.6 | 0.0500 | 0.250 |
| 35.8 | 0.0750 | 0.375 |
| 48.3 | 0.1000 | 0.500 |

**10 nM Substrate**

| Time (sec) | Fraction of Substrate reacted | Concentration of Substrate reacted (nM) |
|---|---|---|
| 0.0 | 0.0000 | 0.000 |
| 12.6 | 0.0250 | 0.250 |
| 25.4 | 0.0500 | 0.500 |
| 38.5 | 0.0750 | 0.750 |
| 51.8 | 0.1000 | 1.000 |

**20 nM Substrate**

| Time (sec) | Fraction of Substrate reacted | Concentration of Substrate reacted (nM) |
|---|---|---|
| 0.0 | 0.0000 | 0.000 |
| 16.8 | 0.0250 | 0.500 |
| 33.8 | 0.0500 | 1.000 |
| 51.0 | 0.0750 | 1.500 |
| 68.5 | 0.1000 | 2.000 |

**50 nM Substrate**

| Time (sec) | Fraction of Substrate reacted | Concentration of Substrate reacted (nM) |
|---|---|---|
| 0.0 | 0.0000 | 0.000 |
| 29.3 | 0.0250 | 1.250 |
| 58.8 | 0.0500 | 2.500 |
| 88.5 | 0.0750 | 3.750 |
| 118.5 | 0.1000 | 5.000 |

**100 nM Substrate**

| Time (sec) | Fraction of Substrate reacted | Concentration of Substrate reacted (nM) |
|---|---|---|
| 0.0 | 0.0000 | 0.00 |
| 50.1 | 0.0250 | 2.50 |
| 100.4 | 0.0500 | 5.00 |
| 151.0 | 0.0750 | 7.50 |
| 201.8 | 0.1000 | 10.00 |

**Table 8 A fictional data set of chemical reaction that follows Michaelis-Menten theory**

**A concentration of enzyme used in this fictional reaction is 0.03 nM.**

This fictional data (Table 8) was created using values of $k_{cat}$ = 2 $s^{-1}$ and $K_M$ = 20 nM and an enzyme concentration of 0.03 nM for an enzyme that follows Michaelis-Menten theory. However, a 10 % error for the enzyme concentration was engineered into the data for 5 nM substrate to demonstrate how the different methods for calculating $k_{cat}$ and $K_M$ may be affected by experimental errors in real data.

### *A Conventional Method for Determining $k_{cat}$ & $K_M$ using Lineweaver-Burk plot*

A graph of the fraction of substrate reacted against time for the fictional data (Table 8) was plotted in Figure 40. As the concentration of enzyme rises, the progress curve becomes steeper, indicating that the rate of reaction is increasing. The velocity of each reaction can be calculated from the slope of these plots.

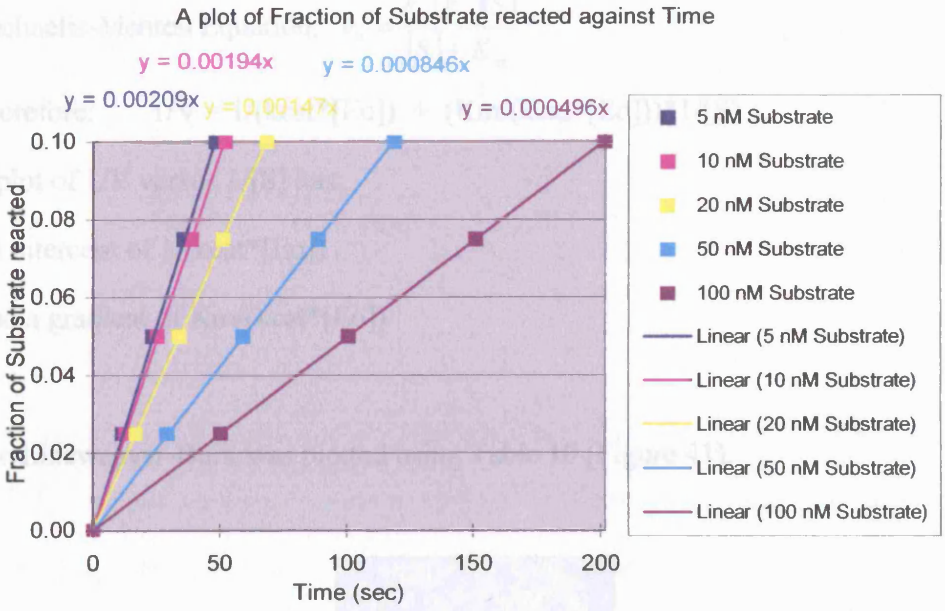



**Figure 40 The linear curves of fraction of substrate reacted over the initial time period**

**Concentrations of substrate are 5, 10, 20, 50 and 100 nM, and enzyme is 0.03 nM.**

The slope of each linear curve (Figure 40) was read as an initial velocity and summarized in Table 9.

| Concentration of Substrate (nM) | Initial Velocity * (fraction reacted per sec.) | Initial Velocity, V (nM per sec.) |
|---|---|---|
| 5 | 0.00209 | 0.01045 |
| 10 | 0.00194 | 0.0194 |
| 20 | 0.00147 | 0.0294 |
| 50 | 0.000846 | 0.0423 |
| 100 | 0.000496 | 0.0496 |

**Table 9 A table of Concentration of Enzyme and Initial Velocity of the fictional data**

*** The initial velocity was obtained from the slope of linear curves in Figure 40.**

$K_m$ and $V_{max}$ can be determined from a Lineweaver-Burk plot of 1/V versus 1/[S] (where V = velocity calculated from the slope of linear curve in Figure 40).

Michaelis-Menten Equation: $v_0 = \frac{k_2[E_0][S]}{[S]+K_m}$

Therefore: $1/V = 1/(kcat*[Eo]) + (Km/(kcat*[Eo]))*1/[S]$

A plot of 1/V versus 1/[S] has:

An intercept of 1/(kcat*[Eo])

And a gradient of Km/(kcat*[Eo])

The Lineweaver-Burk was plotted using Table 10 (Figure 41).

| 1/[S] ($nM^{-1}$) | 1/V ($nM^{-1}$ s) |
|---|---|
| 0.2 | 95.7 |
| 0.1 | 51.5 |
| 0.05 | 34.0 |
| 0.02 | 23.6 |
| 0.01 | 20.2 |

**Table 10 A table of 1/[S] and 1/V**

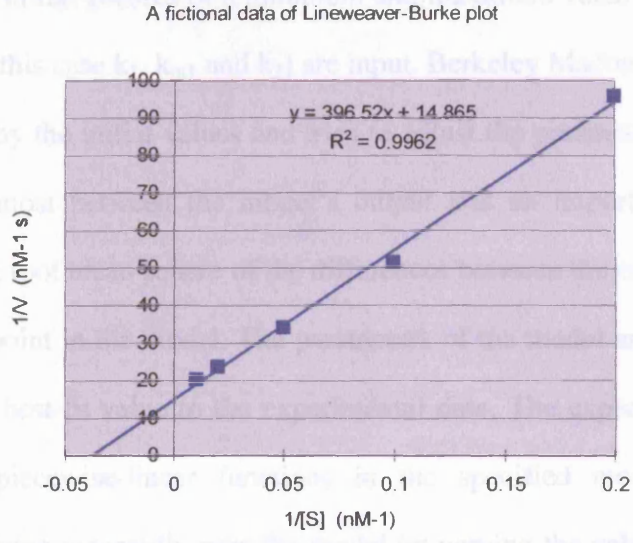



**Figure 41 A Lineweaver-Burk plot of fictional data**

From the Lineweaver-Burk plot, the Y-intercept ($=1/V_{max}$) is 14.87 $nM^{-1}$ s. Therefore $V_{max}$ = 0.0673 nM $^{s-1}$ and since Eo = 0.03 nM, $\mathbf{k_{cat} = 2.24\ s^{-1}}$. The gradient of the Lineweaver-Burk plot is 396.5 s, and this represents $K_M/V_{max}$. Therefore, $\mathbf{K_M = (396.5/14.87) = 26.7\ nM}$.

***Determination of $k_{cat}$ & $K_M$ of the fictional data using Berkeley Madonna***

The differential equation solving programme Berkeley Madonna was used to estimate $k_{cat}$ and $K_M$ of the enzyme for the same set of fictional data (Table 8). The Michaelis-Menten equations were specified in Berkeley Madonna (Figure 42) and the differential equations were solved by the Runge-Kutta 4 method using a fixed-stepsize integration method.

As a first step, initial guesses of a minimum and maximum value for one or more parameters (in this case $k_1$, $k_{m1}$ and $k_2$) are input. Berkeley Madonna generates the curve defined by the initial values and tries to adjust the parameters, to minimise the deviation most between the model's output and an imported dataset. The deviation is the root mean square of the differences between the experimental data point and the point in the model. The parameters of the model are adjusted so as to provide the best-fit value to the experimental data. The experimental datasets are used as piecewise-linear functions in the specified model's equations. Berkeley Madonna repeatedly runs the model by varying the value of parameters until the curve reaches closest to the data points using linear interpolation.

```
METHOD RK4
STARTTIME = 0
STOPTIME = 220
DT = 0.001

E = 0.03
X = 20

INIT Ea = E
INIT Sa = X/4
INIT ESa = 0
INIT Pa = 0
Ma = Pa/(X/4)
d/dt(Ea) = -k1*Ea*Sa + km1*ESa + k2*ESa
d/dt(Sa) = -k1*Ea*Sa + km1*ESa
d/dt(ESa) = k1*Ea*Sa - km1*ESa - k2*ESa
d/dt(Pa) = k2*ESa

INIT Eb = E
INIT Sb = X/2
INIT ESb = 0
INIT Pb = 0
Mb = Pb/(X/2)
d/dt(Eb) = -k1*Eb*Sb + km1*ESb + k2*ESb
d/dt(Sb) = -k1*Eb*Sb + km1*ESb
d/dt(ESb) = k1*Eb*Sb - km1*ESb - k2*ESb
d/dt(Pb) = k2*ESb

INIT Ec = E
INIT Sc = X
INIT ESc = 0
INIT Pc = 0
Mc = Pc/(X)
d/dt(Ec) = -k1*Ec*Sc + km1*ESc + k2*ESc
d/dt(Sc) = -k1*Ec*Sc + km1*ESc
d/dt(ESc) = k1*Ec*Sc - km1*ESc - k2*ESc
d/dt(Pc) = k2*ESc

INIT Ed = E
INIT Sd = 2.5*X
INIT ESd = 0
INIT Pd = 0
Md = Pd/(2.5*X)
d/dt(Ed) = -k1*Ed*Sd + km1*ESd + k2*ESd
d/dt(Sd) = -k1*Ed*Sd + km1*ESd
d/dt(ESd) = k1*Ed*Sd - km1*ESd - k2*ESd
d/dt(Pd) = k2*ESd

INIT Ee = E
INIT Se = 5*X
INIT ESe = 0
INIT Pe = 0
Me = Pe/(5*X)
d/dt(Ee) = -k1*Ee*Se + km1*ESe + k2*ESe
d/dt(Se) = -k1*Ee*Se + km1*ESe
d/dt(ESe) = k1*Ee*Se - km1*ESe - k2*ESe
d/dt(Pe) = k2*ESe

k1 = 10
km1 = 200
k2 = 2
```

**Figure 42 Michaelis-Menten Equations specified in Berkeley Madonna**

Note that the above equations allow Madonna to fit all of the data for all of the substrate concentrations simultaneously. The result of Berkeley Madonna displayed that the calculated curves were almost same as actual data points (the graph is not shown) with the values of parameters;

**$k_{cat}$ = 2.00 $s^{-1}$ and $K_M$ = 19.8 nM.**

*Summary*

The results of two different methods were summarised in Table 11.

| | $k_{cat}$ ($s^{-1}$) | $K_M$ (nM) = $k_{-1}/k_1$ |
|---|---|---|
| **Lineweaver-Burk** | **2.24** | **26.7** |
| **Berkeley Madonna** | **2.00** | **19.8** |

**Table 11 Summary of $k_{cat}$ and $K_M$ obtained by two different methods; the Lineweaver-Burk plot and Berkeley Madonna**

The results indicated that the two different methods reached approximately equivalent values of $k_{cat}$ and $K_M$. The outputs of Berkeley Madonna are much closer to the theoretical values ($k_{cat}$ = 2 $s^{-1}$ and $K_M$ = 20 nM) than those of Lineweaver-Burk plot. This is due to the Lineweaver-Burk plot being influenced by the error created in the data of 5 nM. Because the Lineweaver-Burk is a plot of the reciprocal of the substrate concentration, it is particularly susceptible to errors at low substrate concentrations. In contrast, Berkeley Madonna is much less

influenced by one diverted data set since it attempts to resolve best values to fit *all* of the data at the same time.

Therefore, Berkeley Madonna is a useful tool that is comparable, and in some circumstances superior to, conventional plotting methods for calculating rate constants in enzyme reactions. However, it is particularly useful for reaction schemes, such as that of TDG, that have rate equations that cannot be solved. In these cases, rate parameters cannot be determined using conventional linear plotting techniques.

## 4.3.2. Optimising Conditions for Measuring $K_d$ and $k_2$ Kinetic Constants

Initial experiments to determine $K_d$ and $k_2$ for the single turnover reaction of TDG with the CpG·εC substrate in 50 mM KCl found that $K_d$ was too small (< 0.01 nM) to measure under these reaction conditions. The Berthold HPLC radioactivity monitor can accurately measure down to 3 pmol of $^{32}P$ labelled oligonucleotide and this corresponds to 0.3 ml of 0.01 nM; the maximum volume that can sensibly be loaded onto the POROS HQ column. Therefore, 0.01 nM is the lowest concentration of oligonucleotide that can be measured using this glycosylase assay. Negatively-charged DNA is surrounded by positively-charged counter ions. When a protein binds DNA some of these counter ions are displaced. Increasing the concentration of positive ions in the mixture lowers the binding of most DNA-binding proteins through, it is thought, simple competition between counter ions

and protein for binding sites on the DNA (Hard and Lundback, 1996). Increasing the KCl concentration to 150 mM did indeed increase $K_d$ for the TDG reaction with both CpG·T and CpG·εC oligonucleotides. However, because $K_d$ was so much larger for CpG·T than for CpG·εC, the $K_d$ for CpG·T was now so large that prohibitively large quantities of oligonucleotide and TDG would be needed to measure it accurately. Therefore, a concentration of KCl, which allowed $K_d$ for thymine and ethenocytosine substrates to be measured under the same reaction conditions, was needed. This was quite difficult because too much KCl made $K_d$ for CpG·T too large to measure and too little KCl made $K_d$ for CpG·εC smaller than is detectible by this assay. A KCl concentration of 140 mM was found, by trial and error, to be suitable for measuring the $K_d$ of both substrates.

In order to determine accurately $K_d$ and $k_2$ values, reaction rates must be measured over several substrate (and enzyme) concentrations ranging from about $0.1 \times K_d$ to about $10 \times K_d$. In addition, for each concentration samples must be removed at the right time intervals to give an accurate measure of the rate. Rough $K_d$ and $k_2$ values for each substrate were measured in preliminary experiments and then used to design subsequent, more accurate, experiments. Normally, five samples were removed for each concentration at intervals equally spread over the first ~30 % of the reaction.

Here is an example of FPLC analysis of TDG reaction (Figure 43). 50 nM of TDG and $^{32}$P-labelled CpG·T were allowed to react, and then quenched at five different times (see Table 12). The radioactive oligonucleotide was separated into two peaks; product (i.e. thymine-excised oligonucleotide) and intact substrate, by an ion-exchange column. The area of peaks was integrated under product and intact substrate peaks.

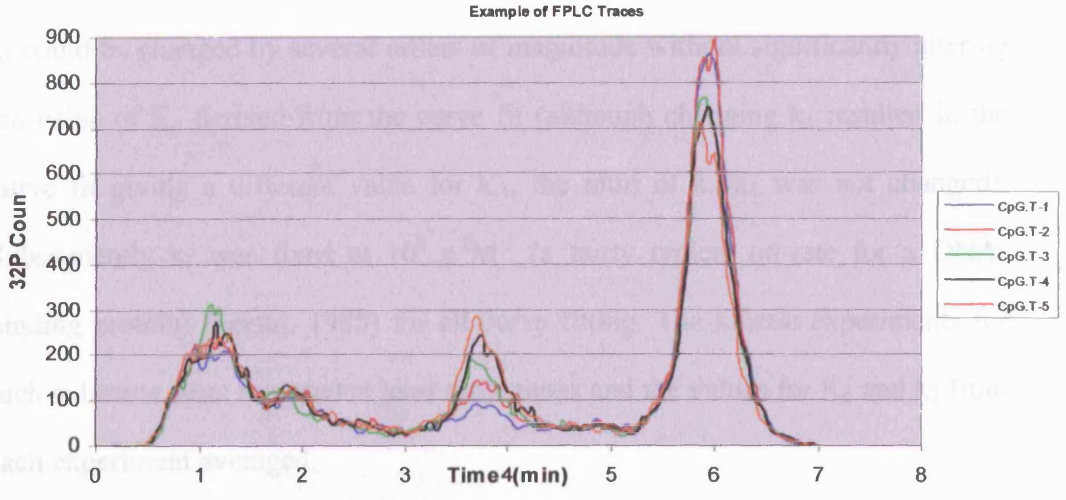



**Figure 43 An example of FPLC analysis of TDG reaction**

**(50 nM of TDG and $^{32}$P-labelled CpG·T; see Table 12), which monitors radioactive count**

**The first peak corresponds to free $^{32}$P that was not incorporated into oligonucleotide during labelling, the second peak indicates a product of TDG reaction (i.e. a fraction of thymine excised) that is eluted earlier than an intact substrate (the third peak). The area was integrated to obtain a fraction of product. The graph shows analysis of the reaction that was quenched at five different times; CpG·T-1 to CpG·T –5.**

## 4.3.3. Determination of $K_d$ and $k_2$ for NpG·T and NpG·εC substrates

Under the determined experimental conditions described above time-course assays were carried out and the kinetic parameters $K_d$ and $k_2$ obtained using the Madonna curve-fitting program. Initially, values of $k_1$, $k_{-1}$, and $k_2$ were allowed to vary in the curve fitting program but this always gave values for $k_1$ that were faster than the diffusion limit ($10^9$ $s^{-1}M^{-1}$) (Fersht, 1985). However, we found that $k_1$ could be changed by several orders of magnitude without significantly altering the value of $K_d$ derived from the curve fit (although changing $k_1$ resulted in the curve fit giving a different value for $k_{-1}$, the ratio of $k_{-1}/k_1$ was not changed). Subsequently $k_1$ was fixed at $10^8$ $s^{-1}M^{-1}$ (a fairly typical on-rate for a DNA-binding protein) (Fersht, 1985) for all curve fitting. The kinetic experiments for each substrate were repeated at least three times and the values for $K_d$ and $k_2$ from each experiment averaged.

Initial concentrations and differential equations for each substance based on the model (Scheme 2) were specified in the Equation Window of Berkeley Madonna, see as an example below:

```
STARTTIME = 0
STOPTIME = 600
DT = 0.1

INT (E) = 25
INT (S) = 25
INT (ES) = 0
INT (EP) = 0
d[E]/dt = -k1*[E]*[S] + k-1*[ES]
d[S]/dt = -k1*[E]*[S] + k-1*[ES]
d[ES]/dt = k1*[E]*[S] - k-1*[ES] - k2*[ES]
d[EP]/dt = k2*[ES]
```

One example of the curve fitting results for determining $K_d$ and $k_2$ of the base excision reaction of TDG with each substrate is given below.

Table 12 - Table 17 were experimental data obtained from time-course assay for each substrate. These data were imported to Berkeley Madonna, and the curve-fitting program output Figure 44 - Figure 49. Some of the data points do not exactly match with calculated lines. This is thought to happen due to pipetting error when one deals with very low concentration of substrate (and enzyme), such as ethenocytosine.

## *Substrate 1: Thymine mismatched with guanine in CpG sequence;* ***CpG·T***

2.5 nM of Substrate/TDG

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 120 | ------ |
| 240 | 0.100 |
| 360 | 0.121 |
| 480 | 0.155 |
| 600 | 0.177 |

5 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 60 | ------ |
| 120 | 0.097 |
| 180 | 0.119 |
| 240 | 0.156 |
| 300 | 0.175 |

12.5 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 34 | 0.097 |
| 64 | 0.113 |
| 96 | 0.153 |
| 128 | 0.182 |
| 160 | 0.201 |

25 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 28 | 0.096 |
| 56 | 0.133 |
| 84 | 0.180 |
| 112 | 0.203 |
| 140 | 0.248 |

50 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 24 | 0.097 |
| 48 | 0.147 |
| 72 | 0.201 |
| 96 | 0.236 |
| 120 | 0.300 |

125 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 18 | 0.081 |
| 36 | 0.143 |
| 54 | 0.205 |
| 72 | 0.246 |
| 90 | 0.327 |

250 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 17 | 0.094 |
| 34 | 0.170 |
| 51 | 0.228 |
| 68 | 0.314 |
| 85 | 0.335 |

**Table 12 The experimental datasets of excision of mismatched thymine in the CpG sequence**

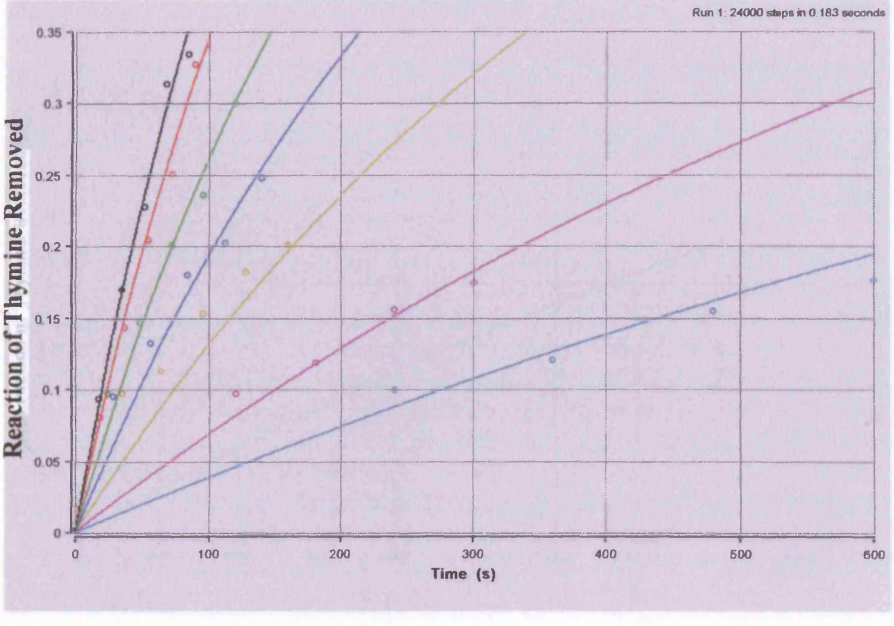



**Figure 44 Time course assay of TDG reaction with substrate; CpG·T**

**Seven different equimolar concentrations (2.5 to 250 nM) of CpG·T and TDG were studied for the rate of single-turnover thymine excision. There were five samples quenching for each concentration of the reaction. The data for all seven different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

## *Substrate 2: Thymine mismatched with guanine in TpG sequence;* ***TpG·T***

5 nM of Substrate/TDG

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 4000 | 0.078 |
| 8000 | 0.138 |
| 12000 | 0.181 |
| 16000 | 0.233 |
| 20004 | 0.258 |

12.5 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 2400 | 0.099 |
| 4800 | 0.157 |
| 7200 | 0.217 |
| 9600 | 0.271 |
| 12000 | 0.331 |

25 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 1600 | 0.091 |
| 3200 | 0.152 |
| 4860 | 0.217 |
| 6400 | 0.266 |
| 8000 | 0.318 |

50 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 960 | 0.081 |
| 1920 | 0.145 |
| 2880 | 0.190 |
| 3840 | 0.239 |
| 4800 | 0.271 |

125 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 720 | 0.098 |
| 1440 | 0.154 |
| 2160 | 0.206 |
| 2880 | 0.259 |
| 3600 | 0.306 |

250 nM

| Time (sec) | Fraction of Thymine removed |
|---|---|
| 680 | 0.093 |
| 1360 | 0.159 |
| 2040 | 0.210 |
| 2720 | 0.275 |
| 3400 | 0.304 |

**Table 13 The experimental datasets of excision of mismatched thymine in the TpG sequence**

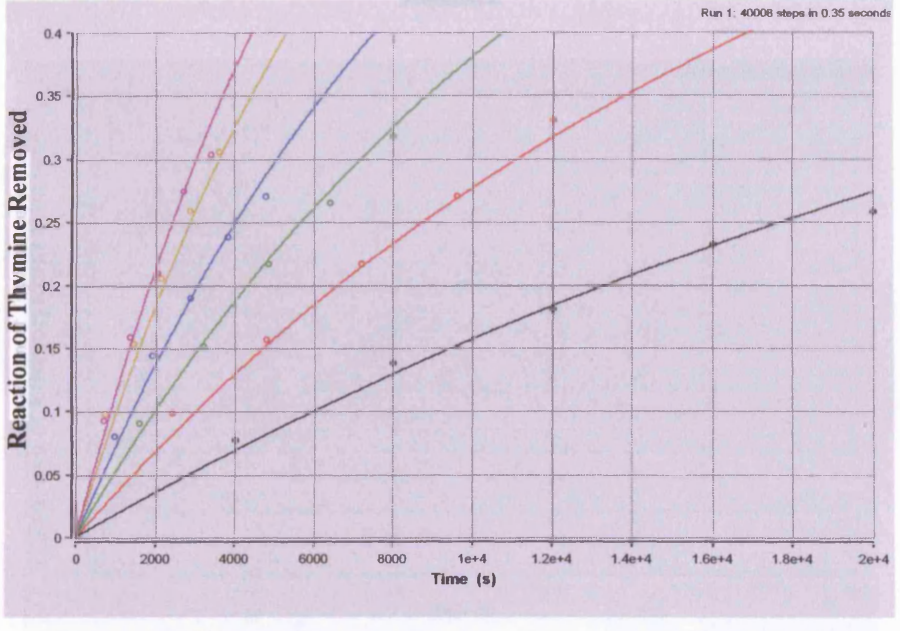



**Figure 45 Time course assay of TDG reaction with substrate; TpG·T**

**Six different equimolar concentrations (5 to 250 nM) of TpG·T and TDG were studied for the rate of single-turnover thymine excision. There were five samples for each concentration of the reaction. The data for all six different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

## *Substrate 3: Ethenocytosine mismatched with guanine in CpG sequence;* ***CpG·εC***

0.01 nM of Substrate/TDG

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 450 | 0.143 |
| 900 | 0.189 |
| 1350 | 0.190 |
| 1800 | 0.213 |
| 2250 | 0.244 |

0.02 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 260 | 0.150 |
| 520 | 0.247 |
| 780 | 0.306 |
| 1040 | 0.355 |
| 1300 | 0.399 |

0.05 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 160 | 0.176 |
| 320 | 0.188 |
| 482 | 0.234 |
| 640 | 0.278 |
| 800 | 0.306 |

0.1 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 120 | 0.138 |
| 240 | 0.211 |
| 360 | 0.263 |
| 480 | 0.321 |
| 600 | 0.344 |

0.2 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 100 | 0.113 |
| 200 | 0.177 |
| 300 | 0.239 |
| 400 | 0.296 |
| 500 | 0.300 |

0.5 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 90 | 0.158 |
| 180 | 0.188 |
| 270 | 0.247 |
| 360 | 0.285 |
| 450 | 0.322 |

1 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 80 | 0.093 |
| 160 | 0.180 |
| 240 | 0.197 |
| 320 | 0.284 |
| 400 | 0.301 |

**Table 14 The experimental datasets of excision of mismatched ethenocytosine in the CpG sequence**

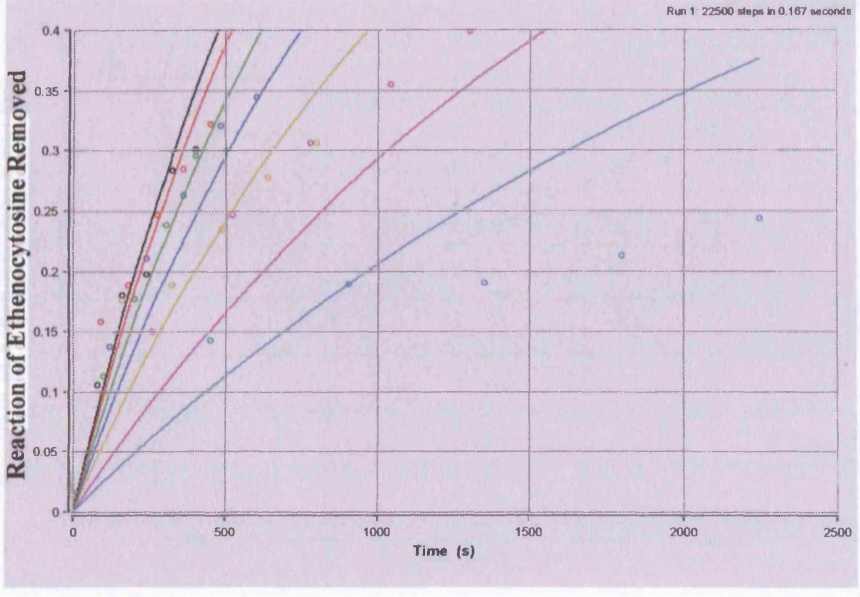



**Figure 46 Time course assay of TDG reaction with substrate; CpG·εC**

**Seven different equimolar concentrations (0.01 to 1 nM) of CpG·εC and TDG were studied for the rate of single-turnover ethenocytosine excision. There were five samples for each concentration of the reaction. The data for all seven different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

## *Substrate 4: Ethenocytosine mismatched with guanine in TpG sequence;* ***TpG·εC***

0.04 nM of Substrate/TDG

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 5760 | 0.123 |
| 11520 | 0.183 |
| 17280 | 0.268 |
| 23040 | 0.327 |
| 28800 | 0.369 |

0.08 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 3380 | 0.129 |
| 6760 | 0.208 |
| 10140 | 0.286 |
| 13640 | 0.318 |
| 17080 | 0.391 |

0.2 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 1800 | 0.171 |
| 3600 | 0.236 |
| 5400 | 0.296 |
| 7200 | 0.348 |
| 9000 | 0.412 |

0.4 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 1200 | 0.127 |
| 2400 | 0.210 |
| 3600 | 0.268 |
| 4800 | 0.323 |
| 6000 | 0.391 |

0.8 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 900 | -------- |
| 1800 | 0.160 |
| 2700 | 0.194 |
| 3600 | 0.286 |
| 4500 | -------- |

2 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 660 | 0.110 |
| 1320 | 0.140 |
| 1980 | 0.222 |
| 2640 | 0.277 |
| 3300 | 0.305 |

4 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 540 | 0.090 |
| 1080 | 0.143 |
| 1620 | 0.183 |
| 2160 | 0.245 |
| 2700 | 0.281 |

**Table 15 The experimental datasets of excision of mismatched ethenocytosine in the TpG sequence**

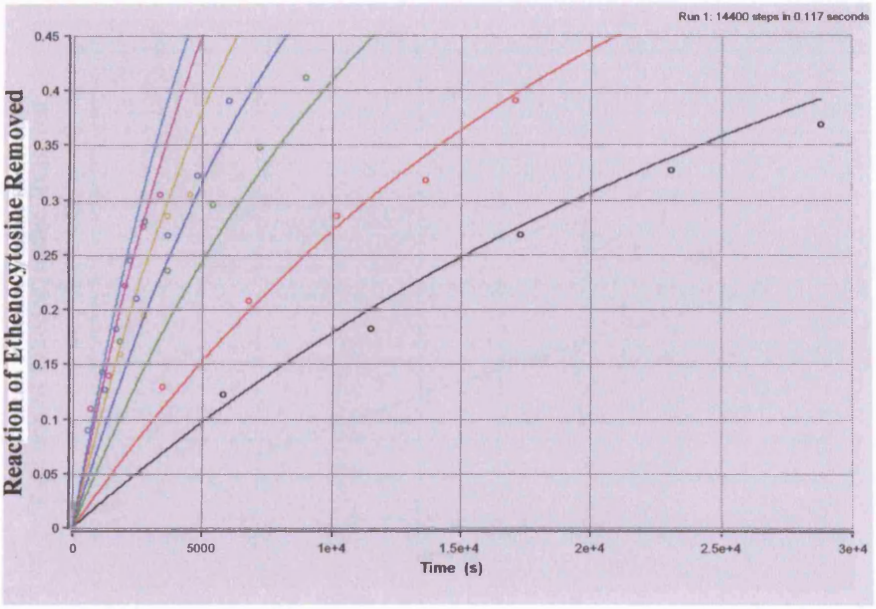



**Figure 47 Time course assay of TDG reaction with substrate; TpG·εC**

**Seven different equimolar concentrations (0.04 to 4 nM) of TpG·εC and TDG were studied for the rate of single-turnover ethenocytosine excision. There were five samples for each concentration of the reaction. The data for all seven different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

## *Substrate 5: Ethenocytosine mismatched with guanine in GpG sequence;* ***GpG·εC***

0.04 nM of Substrate/TDG

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 3390 | 0.085 |
| 6870 | 0.072 |
| 10080 | 0.093 |
| 13560 | 0.093 |
| 16950 | 0.107 |

0.08 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 1950 | 0.098 |
| 3900 | 0.133 |
| 5850 | 0.162 |
| 7800 | 0.213 |
| 9750 | 0.243 |

0.2 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 1020 | 0.141 |
| 2100 | 0.131 |
| 3060 | ------ |
| 4080 | 0.221 |
| 5100 | ------ |

0.4 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 720 | 0.098 |
| 1440 | 0.128 |
| 2160 | 0.137 |
| 2880 | 0.157 |
| 3600 | 0.201 |

0.8 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 540 | 0.079 |
| 1080 | 0.099 |
| 1620 | 0.159 |
| 2160 | 0.184 |
| 2700 | 0.204 |

2 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 420 | 0.061 |
| 840 | 0.072 |
| 1260 | 0.085 |
| 1680 | 0.112 |
| 2100 | 0.125 |

4 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 360 | 0.066 |
| 720 | 0.156 |
| 1080 | 0.196 |
| 1440 | 0.221 |
| 1800 | 0.287 |

**Table 16 The experimental datasets of excision of mismatched ethenocytosine in the GpG sequence**

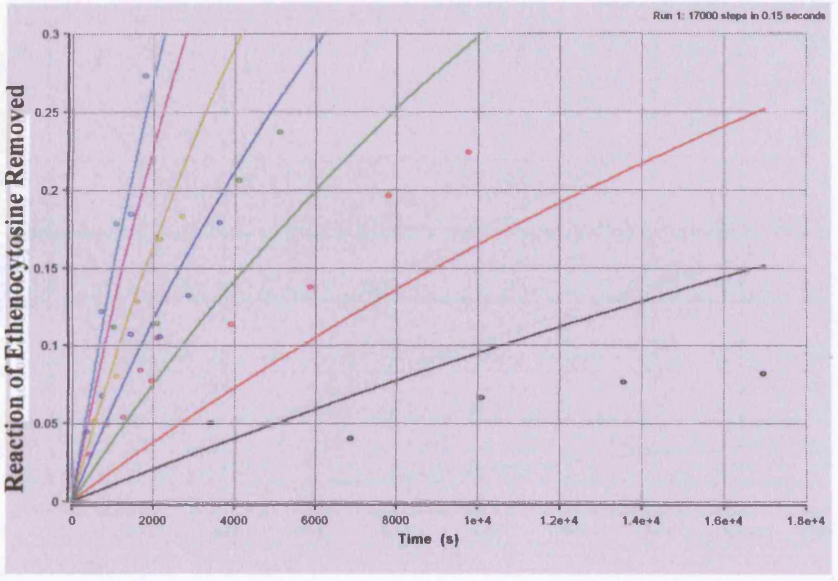



**Figure 48 Time course assay of TDG reaction with substrate; GpG·εC**

**Seven different equimolar concentrations (0.04 to 4 nM) of GpG·εC and TDG were studied for the rate of single-turnover ethenocytosine excision. There were five samples for each concentration of the reaction. The data for all seven different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

## *Substrate 6: Ethenocytosine mismatched with guanine in ApG sequence;* ***ApG·εC***

0.03 nM of Substrate/TDG

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 30000 | 0.017 |
| 66960 | 0.059 |
| 117406 | 0.080 |
| 168960 | 0.110 |

0.06 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 22800 | 0.020 |
| 67020 | 0.074 |
| 82800 | 0.087 |
| 117346 | 0.097 |
| 168900 | 0.119 |

0.15 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 12240 | 0.017 |
| 24000 | 0.030 |
| 47400 | 0.063 |
| 98954 | 0.123 |
| 121214 | ------- |

0.3 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 9120 | 0.033 |
| 19440 | 0.050 |
| 31500 | 0.035 |
| 47280 | 0.146 |
| 98834 | 0.276 |

0.6 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 7500 | 0.037 |
| 15000 | 0.024 |
| 22500 | 0.069 |
| 31680 | 0.117 |
| 47160 | 0.184 |

1.5 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 6000 | 0.029 |
| 12240 | 0.042 |
| 19200 | 0.108 |
| 24000 | 0.135 |
| 45840 | 0.227 |

3 nM

| Time (sec) | Fraction of Ethenocytosine removed |
|---|---|
| 5400 | 0.068 |
| 11880 | 0.103 |
| 16320 | 0.158 |
| 21960 | 0.224 |
| 31200 | 0.278 |

**Table 17 The experimental datasets of excision of mismatched ethenocytosine in the ApG sequence**

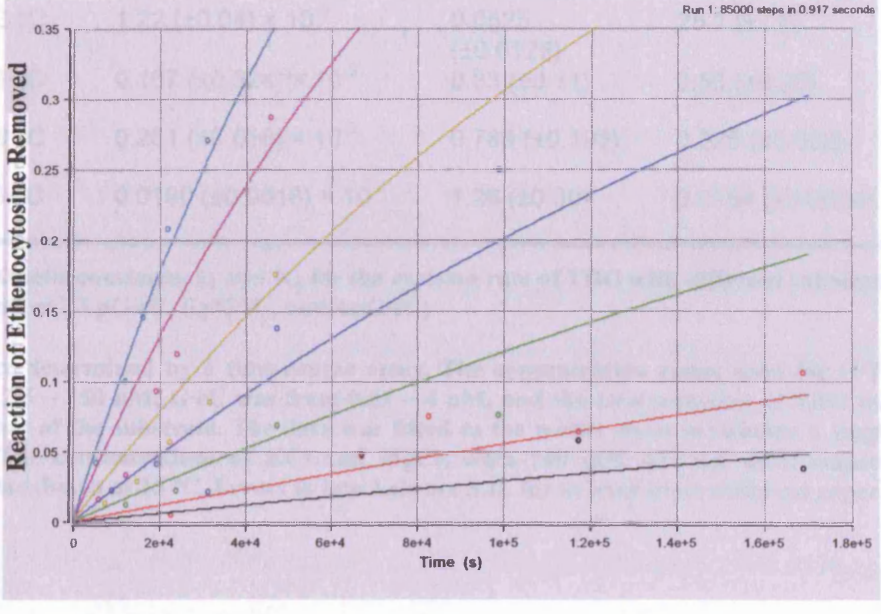



**Figure 49 Time course assay of TDG reaction with substrate; ApG·εC**

**Seven different equimolar concentrations (0.03 to 3 nM) of ApG·εC and TDG were studied for the rate of single-turnover ethenocytosine excision. There were five samples for each concentration of the reaction. The data for all seven different concentration of the reaction were analysed using the curve-fitting Berkeley Madonna software and fitted simultaneously to the reaction scheme given in Scheme 2 to give values for $K_d$ and $k_2$.**

$K_d$ and $k_2$ values were determined by averaging the results of at least three separate experiments. The $K_d$ and $k_2$ for GpG·T and ApG·T substrates could not be attained because the reaction rates were too slow to allow accurate determination under these experimental conditions. The results are given in Table 18 and the significant findings discussed point by point below.

| Substrate | $k_2$ ($s^{-1}$) | $K_d$ (nM) | Specificity constant, $k_2/K_d$ ($\mu M^{-1}s^{-1}$) |
|---|---|---|---|
| CpG·T | 7.71 (±0.38) × $10^{-3}$ | 40.6 (±4.8) | 0.192 (±0.014) |
| TpG·T | 0.118 (±0.056) × $10^{-3}$ | 33.7 (±7.5) | 0.0034 (±0.0010) |
| CpG·εC | 1.22 (±0.04) × $10^{-3}$ | 0.0525 (±0.0125) | 25.1 (±7.6) |
| TpG·εC | 0.167 (±0.024) × $10^{-3}$ | 0.33 (±0.11) | 0.56 (±0.20) |
| GpG·εC | 0.251 (±0.016) × $10^{-3}$ | 0.789 (±0.193) | 0.325 (±0.060) |
| ApG·εC | 0.0190 (±0.0016) × $10^{-3}$ | 1.26 (±0.30) | 0.0154 (±0.0024) |

**Table 18 Kinetic constants, $k_2$ and $K_d$ for the excision rate of TDG with different substrates (CpG·T, TpG·T, CpG·εC, TpG·εC, GpG·εC, and ApG·εC)**

**These were determined by a time-course assay. The concentration range used for G·T mismatch was from 2.5 – 250 nM, G·εC was from 0.01 – 4 nM, and the concentration of TDG was kept the same as that of the substrate. The data was fitted to the model given in Scheme 2 single-turnover reaction. The concentrations of KCl and $MgCl_2$ were 140 mM and 2.5 mM, respectively, and reactions incubated at 25 °C. Errors in brackets are S.D. for at least three different experiments.**

## 4.3.4. CpG·T and CpG·εC; thymine or ethenocytosine in CpG sequence

The results showed that $k_2$ (corresponding to the actual enzymatic step) for CpG·T ($7.71 \times 10^{-3}$ $s^{-1}$) was six times faster than that of CpG·εC ($1.22 \times 10^{-3}$ $s^{-1}$). The $k_2$ for the CpG·T substrate obtained in this experiment and that obtained in (Waters and Swann, 1998) are reasonably similar ($7.71 \times 10^{-3}$ $s^{-1}$ and $15.2 \times 10^{-3}$ $s^{-1}$ respectively). In addition, MED4, which possesses a thymine glycosylase function from G·T mismatches, exhibits a similar value of $12 \times 10^{-3}$ $s^{-1}$ (Petronzelli *et al.*, 2000).

The difference in $K_d$ between CpG·T and CpG·εC was remarkable. The $K_d$ for CpG·εC (0.0525 nM) is nearly 800-fold smaller than that of CpG·T (40.6 nM). This means that TDG binds the ethenocytosine-mismatched substrate more readily than the thymine-mismatched substrate. The binding action of TDG probably involves initial recognition of the mismatched base, followed by flipping of the mismatched base out of the double-stranded DNA into a separate binding site. Therefore $K_d$ must be a consequence not only of how tightly TDG binds the substrate through specific contacts to the mismatched base (and the neighbouring sequence), but also how easily the base is flipped out of the DNA duplex. Flipping involves breaking the hydrogen bonds between the target base and the mismatched guanine. Hence, there are two possible reasons for the small $K_d$ for ethenocytosine substrates. One is that TDG may have a stronger affinity for an ethenocytosine base itself than for a thymine base. The second is that flipping out ethenocytosine is thermodynamically much easier than flipping out thymine. This latter explanation seemed feasible from structural studies of G·T and G·εC mismatches (See Figure 50 and Figure 51)(Allawi and SantaLucia, 1998; Cullinan *et al.*, 1997; Hunter *et al.*, 1987). NMR studies revealed that a DNA duplex containing the G·εC base pair showed a slight bend at the lesion but no major perturbation of the sugar-phosphate backbone (Cullinan *et al.*, 1997). The ethenocytosine is displaced toward the major groove and is poorly stacked with the adjacent bases while the complementary base, guanine is not shifted and remains stacked. This distorted, and unstable structure has one hydrogen bond between the ethenocytosine and the mismatched guanine (Allawi and SantaLucia,

1998; Cullinan *et al.*, 1997). In contrast, thymine in a G·T mismatch is displaced into the major groove, but is well stacked with the adjacent bases and the complementary guanine is pushed toward the minor groove (Hunter *et al.*, 1987). The G·T mismatch is called a "wobble" base pair and is quite stable having two hydrogen bonds between the guanine and the thymine. Thermodynamic studies showed that a 15-mer duplex containing a G·εC mismatch lowered the melting temperature relative to the control duplex (containing a G·C base pair) by 11.6 ℃ whereas a G·T mismatch destabilised by only 7.2 ℃ (Sagi *et al.*, 2000). Another group found that a single G·εC in a 13-mer duplex induces a thermodynamic destabilization (ΔΔG) of –13.8 kcal/mol (Gelfand *et al.*, 1998). It is therefore quite possible that the less stable G·εC base pair can be kinked and the ethenocytosine flipped out by TDG more efficiently than it can flip out thymine from a G·T mismatch (Sagi *et al.*, 2000).

To examine which factor caused the large difference in $K_d$ between CpG·T and CpG·εC free ethenocytosine and thymine bases were used as inhibitors (see Table 19 and Figure 52). If the rate of excision is inhibited significantly more by free ethenocytosine in the reaction mixture than by free thymine, one can deduce that the small $K_d$ of ethenocytosine-mismatched substrate comes from TDG making better contact with the flipped out ethenocytosine base itself than with flipped out thymine.

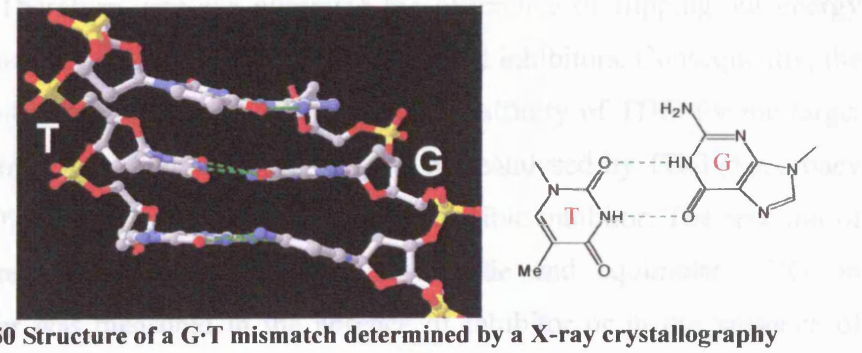



**Figure 50 Structure of a G·T mismatch determined by a X-ray crystallography**

**The figure was taken from Hunter *et al.* (1987).**

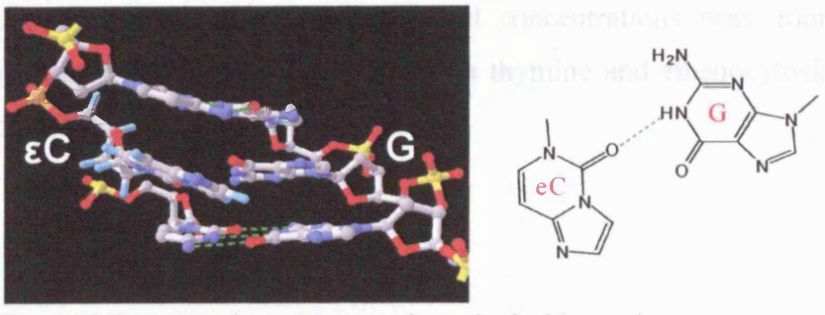



**Figure 51 Structure of an ethenocytosine paired with guanine**

**The figure was taken from an NMR study by Cullinan *et al.* (1997). These structural studies (Figure 50 and Figure 51) revealed that the G·εC base pair is distorted that causes unstable structure if compared with the structure of G·T mismatch pair that integrates well in the double stranded DNA. It shows that the G·εC base pair only has one hydrogen bond between them while the G·T pair manages to have two hydrogen bonds that guarantee more thermodynamic stability.**

This inhibition experiment was performed with 0.1 nM of fluorescently labelled CpG·T oligonucleotide and TDG in reaction buffer in the presence of no inhibitor, ethenocytosine base or thymine base. No inhibition was observed in the presence of up to 5 mM of either of the bases (result now shown). One possible reason for this lack of inhibition is that TDG does not have a strong enough affinity for the bases and may require additional contacts with the backbone of DNA.

Subsequently, another inhibition experiment was attempted using a single-stranded oligonucleotide inhibitor that contains either an ethenocytosine or

thymine at the same position. The single-stranded oligonucleotide does not have a rigid structure and there are no base-pairing hydrogen bonds or base stacking interactions. Therefore, one can eliminate the difference of flipping-out energy between ethenocytosine and thymine single-stranded inhibitors. Consequently, the degree of inhibition should only depend upon the affinity of TDG for the target base. It is known that single-stranded DNA is not catalysed by TDG (Saparbaev and Laval, 1998) and therefore behaves as a reversible inhibitor. The reaction of 25 nM fluorescent-labelled CpG·T oligonucleotide and equimolar TDG, in reaction buffer was measured in the absence of inhibitor or in the presence of either single-stranded oligonucleotide containing ethenocytosine or single-stranded oligonucleotide containing thymine. The experiments were repeated with several different concentrations of inhibitors until concentrations were found which highlighted the degree of inhibition between thymine and ethenocytosine single-stranded inhibitors (Table 19 and Figure 52).

| Inhibitor; None (Control) | |
|---|---|
| Time (min) | Fraction of Thymine excised |
| 3 | 0.138 |
| 5 | 0.204 |
| 6.5 | 0.256 |
| 8 | 0.292 |
| 10 | 0.358 |

| Inhibitor; 250 nM s.s. Oligo with T | |
|---|---|
| Time (min) | Fraction of Thymine excised |
| 3 | 0.080 |
| 6 | 0.128 |
| 10 | 0.187 |
| 15 | 0.237 |
| 20 | 0.298 |

| Inhibitor; 60 nM s.s. Oligo with εC | |
|---|---|
| Time (min) | Fraction of Thymine excised |
| 3 | 0.101 |
| 6 | 0.167 |
| 10 | 0.243 |
| 15 | 0.299 |
| 20 | 0.356 |

**Table 19 The Effect of Inhibition of TDG reaction with single-stranded oligonucleotides**

**25 nM of fluorescent-labelled CpG·T mismatch oligonucleotide and 25 nM TDG were incubated with each inhibitor.**

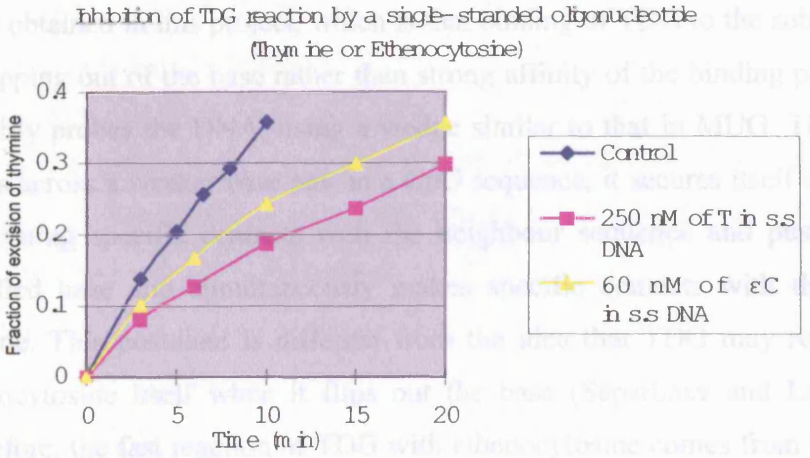



**Figure 52 Inhibition experiment by a single-stranded oligonucleotide that contains either thymine or ethenocytosine at the same position**

**This graph was plotted from the data, Table 19.**

The rate of reaction with 25 nM of substrate was inhibited approximately 50% by 250 nM single-stranded oligonucleotide that contains thymine, and approximately 35% by 60 nM single-stranded oligonucleotide that contains ethenocytosine. In other words, the single-stranded oligonucleotide that contains ethenocytosine inhibited the rate of TDG reaction three times more than single-stranded oligonucleotide that contains thymine. This inhibition experiment indicates that TDG has approximately three-fold greater affinity for ethenocytosine than for thymine and so affinity for the flipped out base cannot be the main factor responsible for the 800-fold difference of $K_d$ between CpG·T and CpG·εC. Thus, the 800-fold difference of $K_d$ most likely comes from TDG requiring much less energy to flip out ethenocytosine from a CpG·εC base pair than to flip out thymine from a CpG·T mismatch.

Based on a crystal structure, Barrett, *et al.* (1998) suggested that the bacterial homologue of TDG, MUG, pushes out the base by insertion of a "wedge" (Gly143, Leu144, and Arg146) rather than by pulling the base into the binding pocket via strong affinity (Barrett *et al.*, 1998). This assumption agrees with the result obtained in this project, which is that binding of TDG to the substrate relies on flipping out of the base rather than strong affinity of the binding pocket. TDG probably probes the DNA, using a wedge similar to that in MUG. Then when it comes across a weaker base pair in a CpG sequence, it secures itself on the DNA by forming specific contacts with the neighbour sequence and pushes out the modified base and simultaneously makes specific contacts with the widowed guanine. This postulate is different from the idea that TDG may recognize the ethenocytosine itself when it flips out the base (Saparbaev and Laval, 1998). Therefore, the fast reaction of TDG with ethenocytosine comes from the unstable structure of G·εC base pair rather than the specificity of the binding pocket for the ethenocytosine itself. Once ethenocytosine or thymine at a CpG site is flipped out, the glycosidic bond of thymine is cleaved faster than that of ethenocytosine.

## 4.3.5. CpG·T and TpG·T; effect of the base pair 5' to the mismatched guanine on thymine excision

The difference in specificity constant between CpG·T and TpG·T was nearly 60-fold. This means that TDG must make specific contacts with the 5' C·G pair of a CpG·T mismatch. The large difference in specificity comes from CpG·T having a greater $k_2$ with virtually no difference in $K_d$ and so this contact seems not to have any effect on formation of the enzyme-substrate complex. Instead it contributes to efficient TDG catalysis by stabilizing the transition-state and reducing the activation energy. This agrees with a methylation interference study indicating that TDG has a strong contact with the 7-nitrogen of the guanine in the 5' C·G base pair (Scharer *et al.*, 1997). Furthermore, as Waters and Swann (1998) and Barrett *et al.* (1998; 1999) have suggested, the N-1 hydrogen of the mismatched guanine might form a hydrogen bond with the TDG peptide inserted into the DNA. Waters *et al.* have also implied that the contacts that TDG makes to the 5'-C·G base pair and the mismatched guanine are cooperative (Waters and Swann, 1998). As a sequence alignment of TDG with *E.coli* homologue MUG shows (Figure 53), a region unique to TDG (AAs 276 – 289 shown in pink in the alignment) might play an important role for recognition of 5' C·G base pair. Because the region starts right after a "wedge" region that might have a specific contact with the mismatched guanine, which is implying that the unique region might lie near by the 5'-C·G base pair.

This sequence dependence of the TDG reaction strongly implies that the enzyme has evolved to repair the G·T mismatches that arise from 5-methylcytosine deamination. Because cytosine is only methylated at CpG sequences deamination would produce a CpG·T mismatch.

**Human TDG and *E. coli* MUG Alignment**

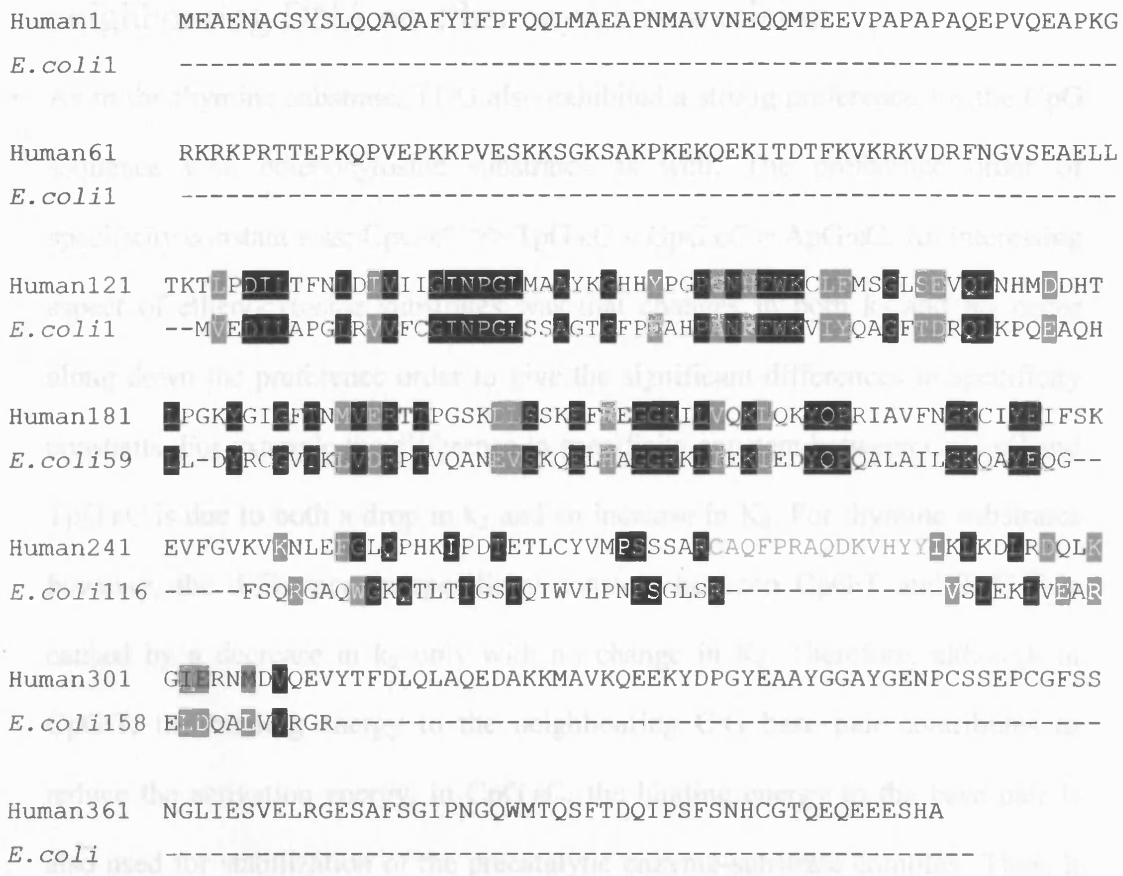



**Figure 53 Amino acid sequence alignment of human thymine-DNA glycosylase with *E.coli* mismatch-specific uracil DNA Glycosylase**

**Amino acids in pink are unique to human TDG and a difficult region to align with *E. coli* MUG. The region is right after a "wedge" region, therefore it might be important region for recognition of a base pair 5' to the mismatched guanine.**

## 4.3.6. CpG·εC, TpG·εC, GpG·εC, and ApG·εC; Effect of neighbouring DNA on ethenocytosine excision

As in the thymine substrate, TDG also exhibited a strong preference for the CpG sequence with ethenocytosine substrates as well. The preference order of specificity constant was; CpG·εC >> TpG·εC ≈ GpG·εC > ApG·εC. An interesting aspect of ethenocytosine substrates was that changes in both $k_2$ and $K_d$ occur along down the preference order to give the significant differences in specificity constants. For example the difference in specificity constant between CpG·εC and TpG·εC is due to both a drop in $k_2$ and an increase in $K_d$. For thymine substrates however, the difference in specificity constant between CpG·T and TpG·T is caused by a decrease in $k_2$ only with no change in $K_d$. Therefore, although in CpG·T, the binding energy to the neighbouring C·G base pair contributes to reduce the activation energy, in CpG·εC, the binding energy to the base pair is also used for stabilization of the precatalytic enzyme-substrate complex. Thus, it seems that TDG adapts a different strategy to catalyse different substrates in terms of usage of the binding energy.

In summary, the repair of ethenocytosine by TDG shows a strong dependence on the base pair 5' to the mismatched guanine. However, there is no evidence that ethenocytosine occurs predominantly in the CpG context, and one would expect ethenocytosine to be formed in all sequence contexts. Therefore, ethenocytosine formed in ApG context would be very poorly repaired by TDG (1600-fold slower specificity constant compared to a CpG site). These results suggest that TDG has not evolved to repair ethenocytosine, and refute the suggestion by Hang *et al.* and Saparbaev *et al.* that the main *in vivo* role of TDG is the repair of ethenocytosine.

# 5. RESULTS and DISCUSSION (2); The Study of Interaction of TDG with APEX1

- Effect of APEX1 on the turnover of TDG with either CpG·T or CpG·εC substrate
- Involvement of oligonucleotide on dissociation of TDG by APEX1
- The Study of Protein-Protein Interaction

Base excision repair is sequentially coordinated by passing the intermediate from one step to next (Wilson and Kunkel, 2000). Each member protein recruits the next enzyme to the damaged site, presumably to avoid exposing the unstable repair intermediate to the environment. However, there is a lack of information on interaction between TDG and APEX1. It is known that APEX1 displaces TDG from its abasic product (Waters *et al.*, 1999), but the mechanism for the displacement is not known. How APEX1 utilizes the DNA for the displacement of TDG was studied in the first part of this section, by using different length oligonucleotides. In the second part, possible protein-protein interactions were investigated using a pull-down assay and isothermal titration calorimetry.

# 5.1. Effect of APEX1 on the turnover of TDG with either CpG·T or CpG·εC substrates

Due to the interesting differences in kinetic results between thymine and ethenocytosine substrates, the effect of APEX1 on displacing TDG from the abasic site was tested with the substrates, CpG·T and CpG·εC.

The base excision rate of 10 nM TDG with 100 nM of ether CpG·T or CpG·εC was measured in reaction buffer (+ 50 mM KCl) in the absence or presence of 20, 40 or 80 nM APEX1 (Table 20/Figure 54 and Table 21/
Figure 55).

| Concentration of APEX1 (nM) | Fraction of Thymine Excised at Each Quenched Time | | | | |
|---|---|---|---|---|---|
| | 10 min | 30 min | 1h.15min | 3h | 5h.30min |
| 0 | 0.100 | 0.131 | 0.166 | 0.168 | 0.168 |
| 20 | 0.084 | 0.171 | 0.183 | 0.222 | 0.268 |
| 40 | 0.084 | 0.145 | 0.196 | 0.250 | 0.319 |
| 80 | 0.067 | 0.118 | 0.210 | 0.283 | 0.394 |

**Table 20 The Experimental data of Excision of Thymine by TDG in the absence/presence of APEX1**

**The reaction mixture contained 100 nM CpG·T substrate, 10 nM TDG and none or 20, 40, 80 nM APEX1, in reaction buffer + 50 mM KCl.**

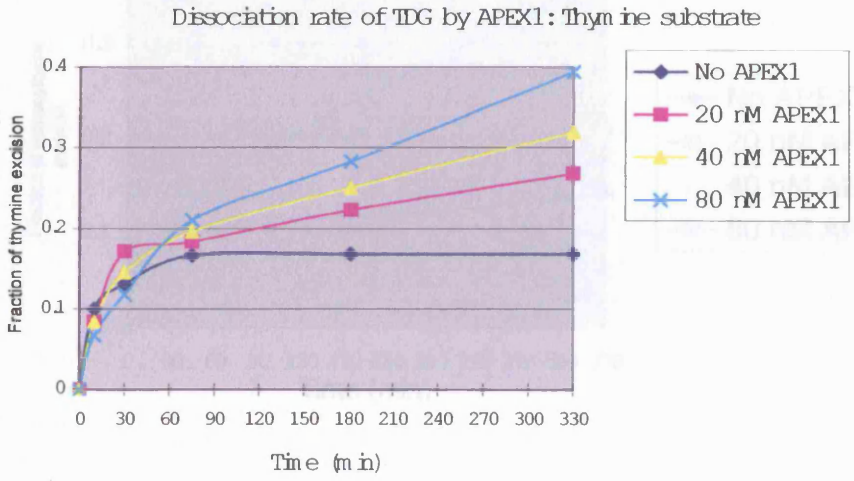



**Figure 54 Turnover of TDG from the product in the absence and the presence of APEX1 using thymine substrate**

**This graph was plotted from the experimental data, Table 20.**

| Concentration of APEX1 (nM) | Fraction of Ethenocytosine Excised at Each Quenched Time | | | | |
|---|---|---|---|---|---|
| | 10 min | 30 min | 1h.15min | 3h | 5h.30min |
| 0 | 0.075 | 0.083 | 0.112 | 0.119 | 0.141 |
| 20 | 0.090 | 0.109 | 0.133 | 0.164 | 0.206 |
| 40 | 0.042 | 0.103 | 0.141 | 0.191 | 0.251 |
| 80 | 0.084 | 0.106 | 0.167 | 0.260 | 0.371 |

**Table 21 The Experimental data of Excision of Ethenocytosine by TDG in the absence/presence of APEX1**

**The reaction mixture contained 100 nM CpG·εC substrate, 10 nM TDG, and none or 20, 40, 80 nM APEX1, in reaction buffer + 50 mM KCl.**

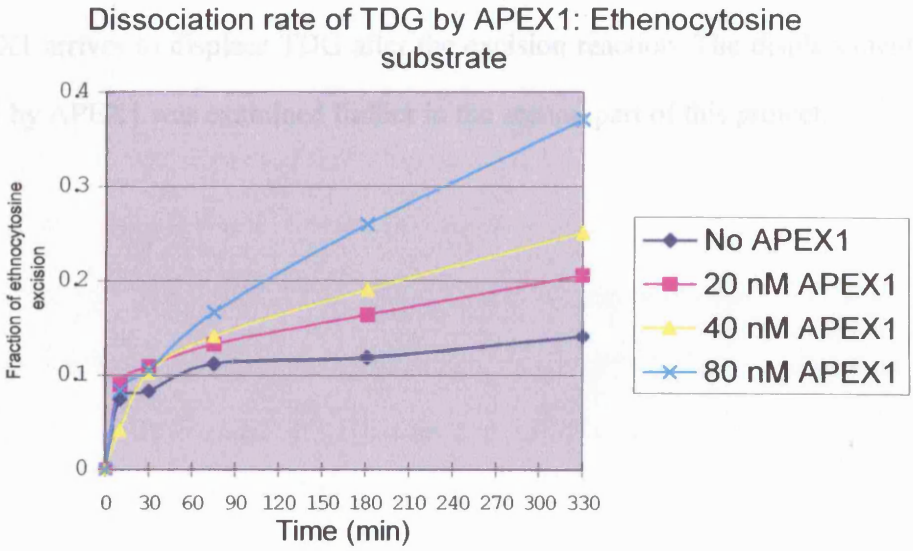



**Figure 55 Turnover of TDG from the product in the absence and the presence of APEX1 using ethenocytosine substrate**

**This graph was plotted from the experimental data, Table 21.**

The effect of APEX1 on increasing the overall turnover of TDG was almost the same between the two substrates. Both results showed that TDG only reacted with a stoichiometric amount of substrate in the absence of APEX1, whereas TDG was released from the product and reacted with more substrate than stoichiometry as the concentration of APEX1 was increased. Thus, as long as an abasic site is present, the origin of the abasic site seems not to be important in terms of the ability of APEX1 to release TDG from the abasic site. This may suggest that the excision reaction of TDG, and the displacement activity and AP endonuclease activity of APEX1 are independent. Thus, it seems that while TDG excises the modified base, APEX1 is not involved in this process, but then APEX1 arrives to displace TDG after the excision reaction. The displacement of TDG by APEX1 was examined further in the second part of this project.

# 5.2. Involvement of oligonucleotide on dissociation of TDG by APEX1

In this study, different lengths of oligonucleotides were used to examine if APEX1 requires the DNA to dissociate TDG from the abasic site. One of the aims of this experiment was to see if one particular side of the DNA was needed for dissociation of TDG by APEX1, and so oligonucleotides were designed to have a G·T mismatch as close to the 5'-end or 3'-end as possible without disrupting a stable complex formation with TDG. The sequences of oligonucleotide used for this experiment are shown in

Table 22. The number in the name of sequence indicates a number of base pairs from a G·T mismatch to each end (E.g. 6G·T18, 6 base pairs and a G·T mismatch then 18 base pairs in the row, in the oligonucleotide).

The DNase I footprint of TDG bound to a non-cleavable oligonucleotide was used as a guide to the length of oligonucleotide required by TDG (Scharer *et al.*, 1997). Also, Wilson *et al.* (1995) found that APEX1 requires four base pairs 5' and three base pairs 3' to the abasic site for catalysis by using an oligonucleotide containing an analogue of an abasic site (tetrahydrofuranyl) at different positions (Willson III *et al.*, 1995).

## Minimum requirement of DNA sequence by TDG and APEX1

```
5'- AGCTTGGCTGCAGGCGGACGGATCCCCGGGAATT  -3'
3'- TCGAACCGACGTCCGTCTGCCTAGGGGCCCTTAA  -5'
```

| Name | | Sequence | |
|---|---|---|---|
| CONTROL (15G·T18) | 5'- | AGCTTGGCTGCAGGCGGACGGATCCCCGGGAATT | -3' (34 Mer) |
| | 3'- | TCGAACCGACGTCCGTCTGCCTAGGGGCCCTTAA | -5' |
| 6G·T18 | 5'- | GCAGGCGGACGGATCCCCGGGAATT | -3' (25 Mer) |
| | 3'- | CGTCCGTCTGCCTAGGGGCCCTTAA | -5' |
| 9G·T18 | 5'- | GCTGCAGGCGGACGGATCCCCGGGAATT | -3' (28 Mer) |
| | 3'- | CGACGTCCGTCTGCCTAGGGGCCCTTAA | -5' |
| 15G·T6 | 5'- | AGCTTGGCTGCAGGCGGACGGA | -3' (22 Mer) |
| | 3'- | TCGAACCGACGTCCGTCTGCCT | -5' |
| 15G·T5 | 5'- | AGCTTGGCTGCAGGCGGACGG | -3' (21 Mer) |
| | 3'- | TCGAACCGACGTCCGTCTGCC | -5' |
| 15G·T4 | 5'- | AGCTTGGCTGCAGGCGGACG | -3' (20 Mer) |
| | 3'- | TCGAACCGACGTCCGTCTGC | -5' |
| MINI (9G·T5) | 5'- | GCTGCAGGCGGACGG | -3' (15 Mer) |
| | 3'- | CGACGTCCGTCTGCC | -5' |

56Mer:

```
5'-CGTAGCTGTACATCAGCTTGGCTGCAGGCGGACGGATCCCCGGGAATTACAGATGC-3'
3'-GCATCGACATGTAGTCGAACCGACGTCCGTCTGCCTAGGGGCCCTTAATGTCTACG-5'
```

**Table 22 List of oligonucleotides used for study of dissociation rate of TDG by APEX1**

**The top sequences above are showing a minimum requirement by TDG and APEX1. The red bases on both sides of G·T mismatch are required by APEX1 (Willson III *et al.*, 1995). The underlined region is bound by TDG (Scharer *et al.*, 1997).**

Initially TDG was incubated with a 5 or 10-fold excess of either control, 56-mer, 15G·T6 or 6G·T18 in the presence of APEX1 to see the effect of APEX1 on dissociation of TDG from its abasic product. The experimental data of effect of APEX1 on dissociation of TDG from its product are shown in Table 23 and a graph was drawn from the table (Figure 56).

| Oligonucleotide | | Quenched time (min) | | | |
|---|---|---|---|---|---|
| | | 35 | 68 | 180 | 300 |
| Control | No APEX1 | 0.194 | 0.198 | 0.196 | 0.199 |
| | 50 nM APEX1 | 0.327 | 0.398 | 0.583 | 0.696 |
| 56Mer | No APEX1 | 0.16 | 0.166 | 0.173 | 0.177 |
| | 50 nM APEX1 | 0.342 | 0.465 | 0.678 | 0.795 |
| 6G·T18 | No APEX1 | 0.079 | 0.095 | 0.128 | 0.155 |
| | 50 nM APEX1 | 0.101 | 0.161 | 0.339 | 0.504 |
| 15G·T6 | No APEX1 | 0.175 | 0.205 | 0.197 | 0.187 |
| | 50 nM APEX1 | 0.196 | 0.239 | 0.322 | 0.4332 |

**Table 23 Fraction of thymine excised by 20 nM of TDG with 100 nM of oligonucleotide (Control, 56Mer, 6G·T18 or 15G·T6) in the absence/presence of APEX1**

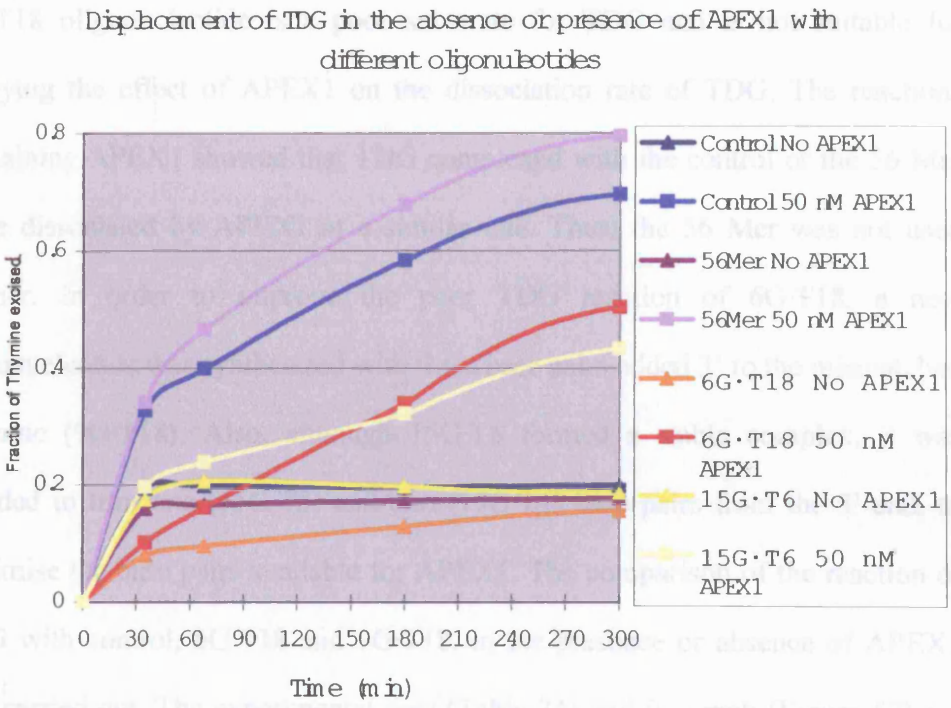



**Figure 56 Reaction of TDG with excess of different oligonucleotides in the absence or presence of APEX1**

**20 nM of TDG and 100 nM of DNA with none or 50 nM of APEX1 were incubated. The reaction was quenched at 35 min, 68 min, 3h, and 5h. The 5'-end of bottom strand was labelled with 6FAM (see Appendix). Substrate and product of TDG were quantified with fluorescent detector.**

Reactions without APEX1 (i.e. only TDG and DNA) were also done to check that each oligonucleotide had the same initial burst rate of reaction as the control followed by the plateau phase when a stoichiometric amount of thymine has been removed. The plateau indicates that TDG can form a stable complex with the abasic product of the oligonucleotide. All the oligonucleotides except 6G·T18 managed to form the stable product complex with TDG (Figure 56). However, the 6G·T18 reaction was much slower and did not achieve the same level of reaction as the other oligonucleotides in the absence of APEX1. Therefore, 6G·T18 oligonucleotide is a poor substrate for TDG and is not suitable for studying the effect of APEX1 on the dissociation rate of TDG. The reactions containing APEX1 showed that TDG complexed with the control or the 56 Mer were dissociated by APEX1 at a similar rate. Thus, the 56 Mer was not used further. In order to improve the poor TDG reaction of 6G·T18, a new oligonucleotide was synthesized with three base pairs added 3' to the mismatched thymine (9G·T18). Also, although 15G·T6 formed a stable complex, it was decided to trim one (15G·T5) and two (15G·T4) base pairs from the 5'-end, to minimise the base pairs available for APEX1. The comparison of the reaction of TDG with control, 6G·T18 and 9G·T18, in the presence or absence of APEX1 was carried out. The experimental data (Table 24) and its graph (Figure 57) are shown below.

| Oligonucleotide | | Quenched time (min) | | | |
|---|---|---|---|---|---|
| | | 30 | 60 | 180 | 300 |
| Control | No APEX1 | 0.151 | 0.122 | 0.141 | 0.151 |
| | 50nM APEX1 | 0.210 | 0.272 | 0.412 | 0.510 |
| 6G·T18 | No APEX1 | 0.050 | 0.064 | 0.113 | 0.152 |
| | 50nM APEX1 | 0.080 | 0.132 | 0.308 | 0.496 |
| 9G·T18 | No APEX1 | 0.129 | 0.145 | 0.147 | 0.147 |
| | 50nM APEX1 | 0.208 | 0.261 | 0.390 | 0.495 |

**Table 24 Fraction of thymine excised by 10 nM of TDG with 100 nM of oligonucleotide (Control, 56Mer, 6G·T18 or 9G·T18) in the absence/presence of APEX1**

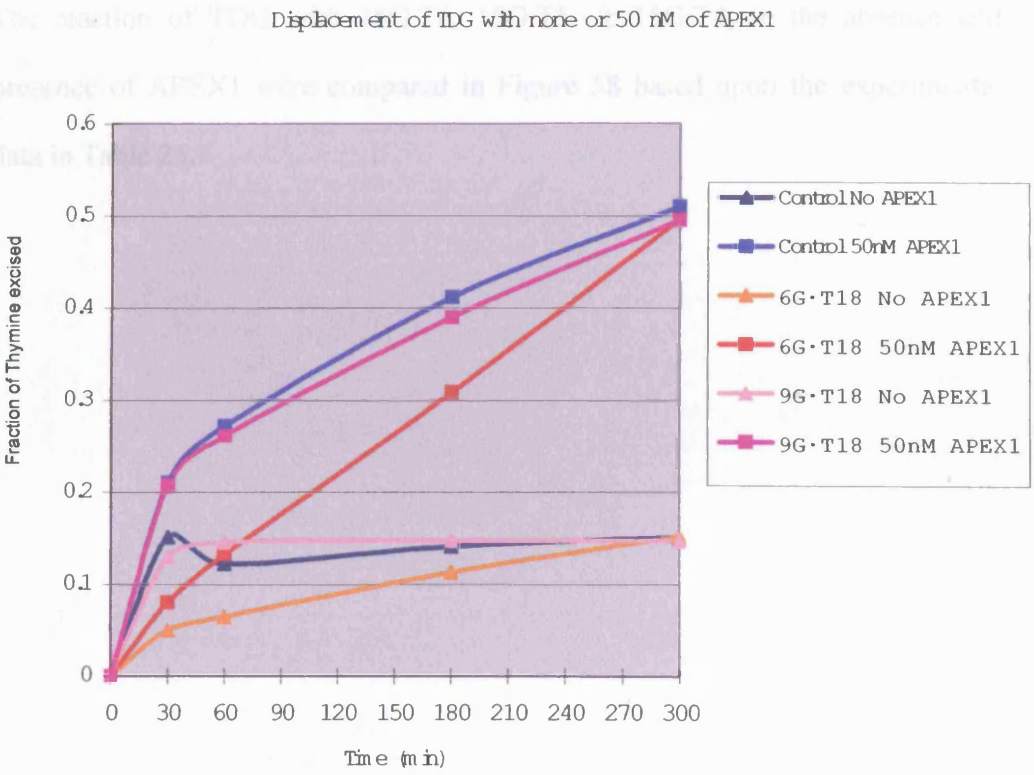



**Figure 57 Effect of APEX1 on dissociation of TDG with varied length of oligonucleotides**

**10 nM of TDG, 100 nM of DNA with none or 50 nM of APEX1 were incubated, and the reaction of TDG was quenched at 30 min, 1h, 3h, and 5h.**

As expected, the 6G·T18 showed exactly the same curves in the absence and presence of APEX1 as seen in Figure 56. However, 9G·T18 displayed the same behaviour as the control oligonucleotide in the absence of APEX1; it had the same initial rate as the control and plateaued after a stoichiometric amount of thymine was removed, showing that it formed a stable product complex with TDG. APEX1 increased the turnover of TDG with 9G·T18 to the same extent as that seen with the control oligonucleotide. The 9G·T18 oligonucleotide thus contains a G·T mismatch as close to the 3'-end of the mismatched thymine that still allows formation of a stable TDG-product complex.

The reaction of TDG with 15G·T6, 15G·T5, & 15G·T4, in the absence and presence of APEX1 were compared in Figure 58 based upon the experimental data in Table 25.

| Oligonucleotide | | Quenched time (min) | | | |
|---|---|---|---|---|---|
| | | 30 | 60 | 200 | 300 |
| Control | No APEX1 | 0.144 | 0.142 | 0.153 | 0.154 |
| | 50nM APEX1 | 0.227 | 0.301 | 0.492 | 0.580 |
| 15G·T6 | No APEX1 | 0.109 | 0.115 | 0.122 | 0.126 |
| | 50nM APEX1 | 0.154 | 0.189 | 0.330 | 0.441 |
| 15G·T5 | No APEX1 | 0.120 | 0.115 | 0.118 | 0.127 |
| | 50nM APEX1 | 0.142 | 0.166 | 0.279 | 0.359 |
| 15G·T4 | No APEX1 | 0.111 | 0.113 | 0.122 | 0.129 |
| | 50nM APEX1 | 0.131 | 0.165 | 0.313 | 0.398 |

**Table 25 Fraction of thymine excised by 10 nM of TDG with 100 nM of oligonucleotide (Control, 15G·T6, 15G·T5 or 15G·T4) in the absence/presence of APEX1**

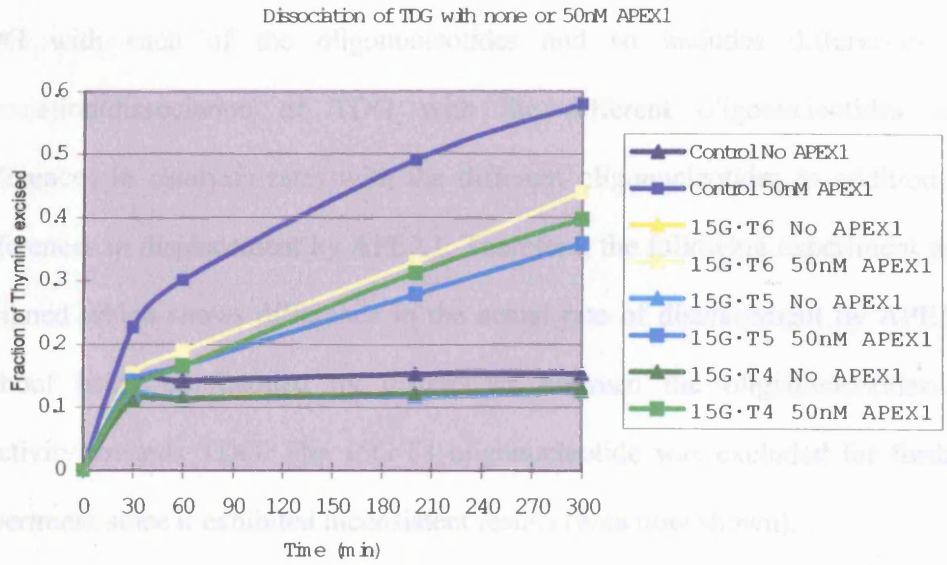



**Figure 58 Comparison of dissociation rate of TDG from varied length of oligonucleotides in the absence and presence of APEX1**

**10 nM of TDG, 100 nM of DNA, and none or 50 nM of APEX were incubated, and the TDG reaction was quenched at 30 min, 1h, 3h20min and 5h.**

All three 5'-truncated oligonucleotides (15G·T6, 15G·T5, and 15G·T4) displayed the same reaction curves as the control oligonucleotide in the absence of APEX1, showing that they were good TDG substrates and formed stable product complexes with TDG. The loss of two further base pairs from the 5'-end seemed not to be important, and all three oligonucleotides showed a similar level of displacement by APEX1. The level of reaction in the presence of APEX1 is significantly higher for the control oligonucleotide than for 15G·T6, 15G·T5, & 15G·T4. This implies that TDG may associate and react slower with 15G·T6, 15G·T5, & 15G·T4 than with the control. However, the steady state rate of dissociation by APEX1 (i.e. slope of the curves after the burst phase) was similar to the control. This experiment looks at the effect of APEX1 on the turnover of TDG with each of the oligonucleotides and so includes differences in association/dissociation of TDG with the different oligonucleotides and differences in catalysis rates with the different oligonucleotides in addition to differences in displacement by APEX1. Therefore, the following experiment was designed which shows difference in the actual rate of displacement by APEX1 without being confounded by differences between the oligonucleotides in reactivity towards TDG. The 15G·T4 oligonucleotide was excluded for further experiment since it exhibited inconsistent results (data now shown).

As a general methodology of next experiment, each different length of oligonucleotide was allowed to form a complex with TDG. Then APEX1 and $^{32}$P-labelled 34 Mer ("control" in Table 22) oligonucleotide were added to the reaction. When APEX displaces TDG from the abasic site the now free TDG will bind to the $^{32}$P-labelled G.T reporter oligonucleotide and remove the thymine

from it. Thus, the amount of thymine excised from the reporter oligonucleotide represents the rate of displacement of TDG by APEX1 The rate of TDG displacement can be determined by fitting a straight line a graph showing thymine removed from the $^{32}$P-labelled reporter oligonucleotide versus time (Figure 59). In the absence of APEX, the rate of thymine excision from the reporter oligonucleotide represents the spontaneous rate of dissociation of TDG from the abasic site. This model makes the reasonable assumption that released TDG reacts with the reporter oligonucleotide much faster than the rate of displacement from the abasic oligonucleotide.

In this experiment, a reaction mixture containing 24 nM of test oligonucleotide and 20 nM of TDG are incubated first, to allow all TDG molecules to react with the oligonucleotide, remove the mismatched thymine and form a complex with the abasic product. Then, 20 nM of $^{32}$P-labelled reporter oligonucleotide (the control sequence), and none or 50 nM of APEX1 were added to the reaction mixture. The results of this experiment with the control oligonucleotide, 9G·T18, 15G·T6 & 15G·T5 and an oligonucleotide truncated of both the 5'- and 3'-end (MINI) are shown in Table 26 and Figure 59. Table 26 shows fraction of thymine excised from the $^{32}$P-reporter oligonucleotide by dissociated TDG.

| Oligonucleotide | | Quenched time (min) | | | | |
|---|---|---|---|---|---|---|
| Control | No APEX1 | 10 | 30 | 60 | 120 | |
| | | 0.0 | 0.037 | 0.047 | 0.053 | |
| | 50nM APEX1 | 3 | 6 | 9 | 12 | 15 |
| | | 0.060 | 0.148 | 0.262 | 0.355 | 0.456 |
| 9G·T18 | No APEX1 | 10 | 30 | 60 | 120 | |
| | | 0.0 | 0.025 | 0.038 | 0.032 | |
| | 50nM APEX1 | 3 | 6 | 9 | 12 | 15 |
| | | 0.048 | 0.124 | 0.207 | 0.290 | 0.366 |
| 15G·T6 | No APEX1 | 10 | 30 | 60 | 120 | |
| | | 0.0 | 0.033 | 0.044 | 0.089 | |
| | 50nM APEX1 | 3 | 6 | 9 | 12 | 15 |
| | | 0.039 | 0.098 | 0.201 | 0.291 | 0.391 |
| 15G·T5 | No APEX1 | 10 | 30 | 60 | 120 | |
| | | 0.0 | 0.030 | 0.042 | 0.062 | |
| | 50nM APEX1 | 3.5 | 7.17 | 11 | 14 | 18 |
| | | 0.037 | 0.100 | 0.172 | 0.236 | 0.338 |
| MINI | No APEX1 | 10 | 30 | 60 | 120 | |
| | | 0.026 | 0.059 | 0.090 | 0.144 | |
| | 50nM APEX1 | 4 | 8 | 12 | 16 | 20 |
| | | 0.037 | 0.081 | 0.136 | 0.222 | 0.283 |

**Table 26 Fraction of thymine excised from 20 nM of $^{32}$P-labelled reporter oligonucleotide by 20 nM of TDG that initially formed complex with 24 nM of each oligonucleotide (Control, 9G·T18, 15G·T6, 15G·T5 or MINI) in the absence/presence of APEX1**

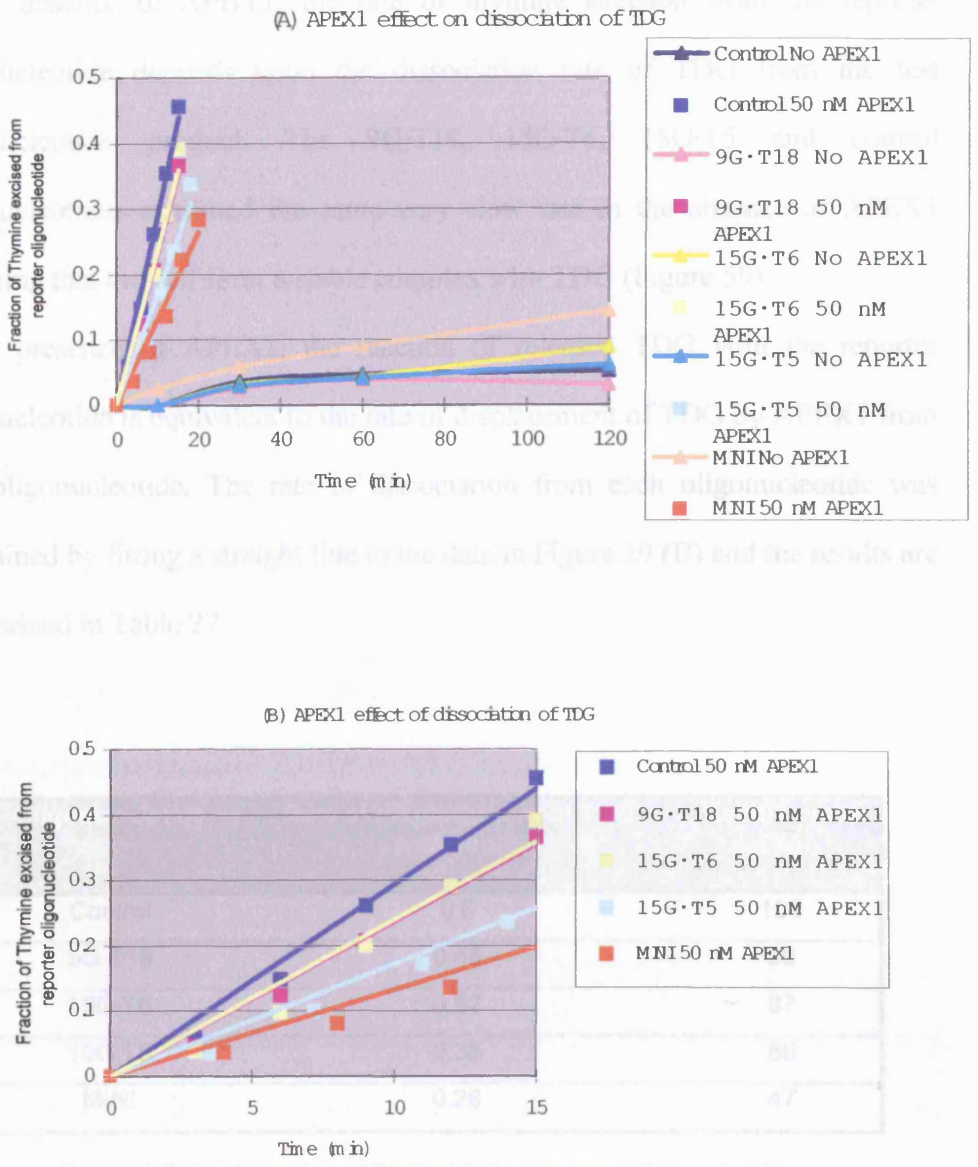



**Figure 59 Effect of DNA on the dissociation of TDG by APEX1**

**(A) 20 nM of TDG was incubated with 24 nM of each oligonucleotide, allowing all the TDG to react and bind to the abasic site. Then, 20 nM of $^{32}$P-labelled reporter oligonucleotide, and none or 50 nM of APEX1 were added to the mixture. Dissociated TDG reacts with the reporter oligonucleotide. The radioactivity of the reaction was analyzed with FPLC. TDG dissociating from the test oligonucleotide reacts with the $^{32}$P-reporter oligonucleotide. The graph shows the extent of excision from this reporter oligonucleotide. (B) The first 15 min of (A) is expanded to highlight the difference of reaction of TDG that was dissociated from each oligonucleotide by APEX1. Only the reactions with 50 nM APEX1 were displayed in B.**

In the absence of APEX1, the rate of thymine excision from the reporter oligonucleotide depends upon the dissociation rate of TDG from the test oligonucleotide product. The 9G·T18, 15G·T6, 15G·T5 and control oligonucleotides exhibited the same very slow rate in the absence of APEX1 indicating that they all form a stable complex with TDG (Figure 59).

In the presence of APEX1, the reaction of released TDG with the reporter oligonucleotide is equivalent to the rate of displacement of TDG by APEX1 from each oligonucleotide. The rate of dissociation from each oligonucleotide was determined by fitting a straight line to the data in Figure 59 (B) and the results are summarized in Table 27.

| Oligonucleotide | Rate of displacement of TDG ($nM\ min^{-1}$) | Relative rate (%) |
|---|---|---|
| Control | 0.6 | 100 |
| 9G·T18 | 0.48 | 80 |
| 15G·T6 | 0.52 | 87 |
| 15G·T5 | 0.36 | 60 |
| MINI | 0.28 | 47 |

**Table 27 Rate of reaction of TDG with the reporter oligonucleotide**

**The TDG (20 nM) was displaced from each test oligonucleotide by 50 nM of APEX1 in order to excise thymine from the reporter oligonucleotide. The displacement rates were calculated from linear fits to the data in Figure 59.**

15G·T6 and 9G·T18 showed similar rate of displacement by APEX1. The 15G·T6 is 12 base pairs shorter (5' to the mismatched T) than the control, and that caused approximately 13 % less dissociation of TDG (Figure 59 and Table 27). Hence,

the 12 base pairs may help for more efficient displacement, but are not an absolute requirement. The 9G·T18 is six base pairs shorter (3' to the mismatched T) than the control, and this caused a similar 20 % drop in displacement rate. Thus, these six base pairs also have only a moderate influence on the displacement of TDG by APEX1.

Interestingly, 15G·T5, which is only one nucleotide shorter than 15G·T6, exhibited approximately 27 % slower displacement than 15G·T6 (approximately 40 % slower than the control). Thus, the sixth base pair 5' to the mismatched T plays a more significant role in the displacement of TDG by APEX1 than all of the other 12 base pairs 5' to it. However, the 15G·T5 still displayed about 60 % displacement rate compared to the control. Thus, this base pair cannot be an absolute requirement of APEX1 for the displacement of TDG.

The MINI oligonucleotide is truncated at both ends and is only 15 base pairs long. From the DNase I footprint, it is expected to be completely covered by TDG, with no free base pairs on either side available for APEX1. The reaction of the reporter oligonucleotide for MINI in the absence of APEX1 was slightly faster than the other oligonucleotides indicating that the length of this oligonucleotide was not quite enough to form as stable a complex with TDG as the control oligonucleotide (although it still dissociates very slowly). Therefore, one has to take into account that some of the TDG has dissociated on its own in the presence of APEX1. The rate of APEX1 displacement of TDG from MINI is about half that of the control oligonucleotide, but has not been completely abolished. Therefore, both the 5'-side and the 3'-side of the DNA contribute to the displacement of TDG, but they are not an absolute requirement of APEX1 for displacement.

*Summary*

In the absence of APEX1, TDG formed very stable product complexes with all of the oligonucleotides, although the complex with MINI dissociated approximately twice as fast. There was a decrease in dissociation of TDG by APEX1 when either side, or both sides of the control oligonucleotide were not available for APEX1; 20 % slower displacement when six base pairs 3' to the mismatched T were removed, 40 % slower displacement when 13 base pairs 5' to the mismatched T were removed, and 50 % slower displacement when both ends were removed (Figure 59 and Table 27). There was no obvious preferred DNA-orientation for APEX1 displacement of TDG. However, there was a trend that the displacement of TDG became less as fewer base pairs were available for APEX1. The MINI oligonucleotide, which should be covered by TDG, exhibited approximately 50 % slower displacement rate than the control. This means that the free DNA on either side of the mismatch may help to load APEX1 onto the TDG-abasic DNA complex. Alternatively, APEX1 might still be able to bind to the DNA on the opposite side to TDG. However, this seems unlikely since there would probably be steric clash between the two enzymes.

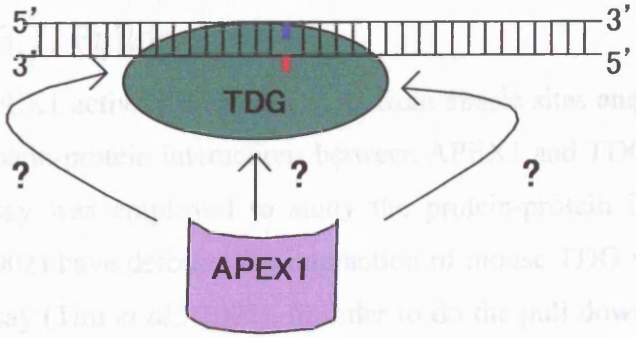
In conclusion, since the displacement activity of APEX1 could not be abolished completely with any oligonucleotides, both sides of DNA from the G·T mismatch are not absolutely required by APEX1 to displace TDG from the abasic site. It can also be concluded that the displacement of TDG by APEX1 does not happen through APEX1 binding and distorting DNA to displace TDG (see Section 2.5). Based on this fact, it can be deduced that APEX1 must have a direct protein-protein interactions with TDG. DNA extending from the TDG complex increases the displacement rate, possibly the DNA might help to guide APEX1 to TDG.

Protein-protein interactions between TDG and APEX1 could be extremely transient and were not detected using a yeast two-hybrid system (Privezentzev *et al.*, 2001). If the protein-protein interaction occurs on the DNA, the interaction of TDG-DNA-APEX1 is also transient since a band-shift experiment failed to detect this ternary complex (Waters *et al.*, 1999). In the next section, detection of protein-protein interactions between APEX1 and TDG was attempted using a pulldown assay and isothermal titration calorimetry.

*<SUMARY of Effect of Oligonucleotides on Displacement of TDG by APEX1>*

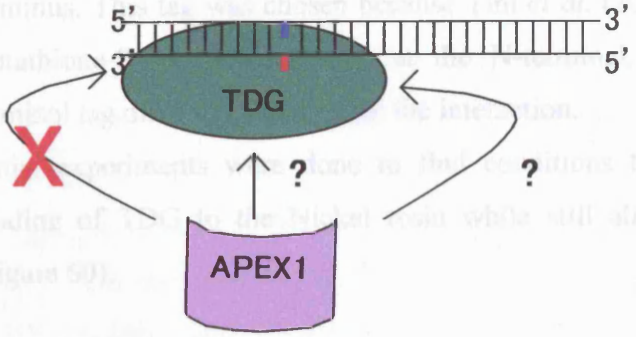
Note: Blue corresponds to guanine, and red is a mismatched thymine in the double-stranded DNA.

Control (34 Mer)



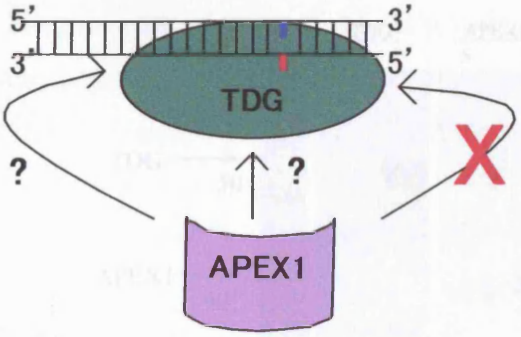

→ The relative rate of displacement; 100 %

6G·T18 (28 Mer)



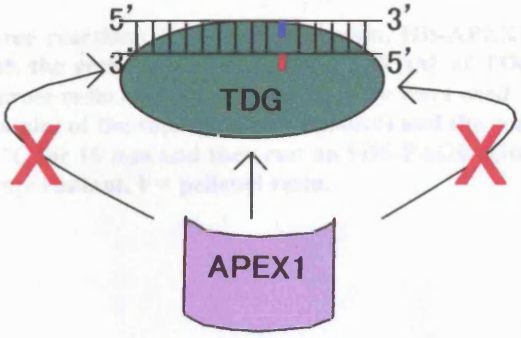

→ The relative rate of displacement; 80 %

15G·T5 (21 Mer)




→ The relative rate of displacement; 60 %

MINI (15 Mer)




→ The relative rate of displacement; 47 %

# 5.3. The Study of Protein-Protein Interaction

## 5.3.1. Pulldown Assay

APEX1 actively displaces TDG from abasic sites and therefore, the possibility of protein-protein interactions between APEX1 and TDG is expected. The pulldown assay was employed to study the protein-protein interaction since Tini *et al.* (2002) have detected the interaction of mouse TDG with APEX1 using the same assay (Tini *et al.*, 2002). In order to do the pull down experiments, the cDNA of APEX1 was subcloned into the pET100 vector to add a six histidine tag to the N-terminus. This tag was chosen because Tini *et al.* (2002) used APEX1 that had a glutathione-S-transferase fusion at the N-terminal, and showed that this N-terminal tag did not interfere with the interaction.

Initial experiments were done to find conditions that eliminated non-specific binding of TDG to the Nickel resin while still allowing His-APEX1 to bind (Figure 60).

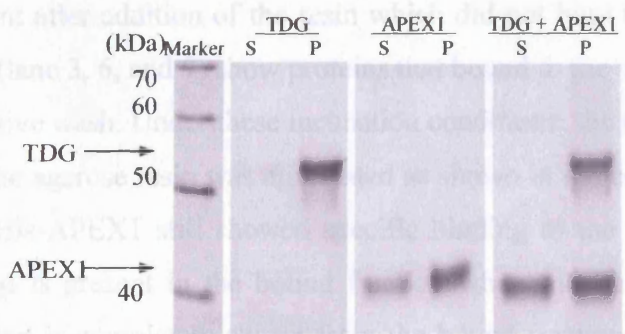



**Figure 60 Pulldown assay of TDG and His-APEX1 showing the non-specific binding of TDG to the nickel-charged agarose resin in the absence of BSA**

**Three reactions, TDG with the resin, His-APEX1 with the resin, and TDG & His-APEX1 with the resin, were carried out. 555 nM of TDG, 493 nM of APEX1, and 1 µl of nickel-agarose resin /15µl of reaction mixture were used. The mixture were incubated for 1h at 4 °C. Samples of the supernatant (unbound) and the washed, pelleted resin (bound) were heated at 90 °C for 10 min and then run on SDS-PAGE. Gels were silver stained to visualize protein. S = supernatant, P = pelleted resin.**

The presence of TDG in the pellet fraction in Figure 60 in the absence of His-APEX1 demonstrates that TDG binds non-specifically to the agarose resin even though the resin was extensively washed before loading onto the SDS-PAGE. The middle panel of Figure 60 shows that, as expected, His-APEX1 binds to the nickel-resin. The bound fraction from the incubation of TDG with His-APEX1 gives a strong band for TDG, suggesting that TDG has been 'pulled-down' by the His-APEX1. However, because TDG binds strongly to the resin in the absence of His-APEX1 one cannot differentiate between non-specific binding of TDG to the resin and specific binding of TDG to His-APEX1. Elimination of non-specific binding of TDG to the resin was attempted by adding BSA and/or imidazole to the reactions. The presence of 27 ng/μl BSA and 20 mM imidazole was found to prevent non-specific binding of TDG to the agarose resin, without stopping His-APEX1 from binding (dissociates from the nickel-charged resin at approximately 80 mM imidazole).

Using these conditions, the pulldown assay was carried out using TDG and His-APEX1 at a concentration ratio of approximately 1:4 (Figure 61). The "Before resin" fractions (lane 1, 4, and 7) show the input of proteins in the reaction mixture. The supernatant fractions (lane 2, 5, and 8) show the proteins in the supernatant after addition of the resin which did not bind to the resin. The pellet fractions (lane 3, 6, and 9) show proteins that bound to the agarose resin even after the extensive wash. Under these incubation conditions, the non-specific binding of TDG to the agarose resin was eliminated as shown in the control of TDG (lane 3) whereas His-APEX1 still showed specific binding to the resin (lane 9). A faint TDG band is present in the bound fraction when incubated together with His-APEX1, but is completely absent from the bound fraction when incubated alone with the resin (Figure 61). This suggests a direct protein-protein interaction of TDG with His-APEX1. The TDG band in the pellet fraction of the TDG-His-APEX1 mixture looks weak on the SDS-PAGE, however, the band consistently appeared in several attempts, and the TDG band was consistently absent from the pellet fraction of TDG and the resin alone (lane 3).

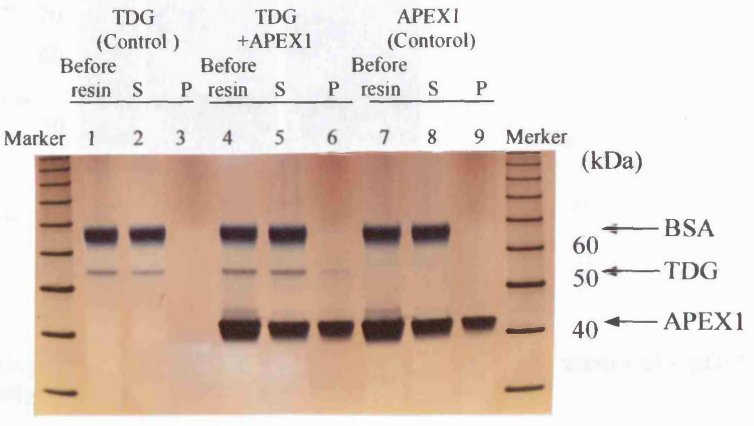



**Figure 61 Pulldown assay; detection of a direct protein-protein interaction between human TDG (390 nM) and APEX1 (1.7 μM)**

**Three reactions, TDG with the resin, APEX1 with the resin, and TDG & APEX1 with the resin, were carried out. A third of the reaction was taken out before the agarose resin was added (lane 1, 4, and 7), another third of the supernatant was taken out after incubation for 1h at 4 °C (lane 2, 5, and 8; S stands for supernatant). The last portion of reaction mixture was spun-down and the pellet extensively washed (lane 3, 6, and 9; P stands for pellet). All samples were heated in SDS-Gel loading buffer for 10 min at 90 °C before loading on the SDS-PAGE.**

Next, the effect of the CpG·T oligonucleotide in the pulldown assay was examined to see if DNA could enhance the interaction between TDG and His-APEX1, possibly via a TDG-DNA-APEX1 ternary complex (Figure 62). However, there was no significant effect of DNA on the amount of TDG pulled down by His-APEX1. This suggest that the DNA is not a prerequisite factor for the protein-protein interaction, and supports the previous experiments with different length oligonucleotides (Section 4.4.1.) that showed that DNA did not have a large effect on the displacement of TDG from abasic sites by APEX1.

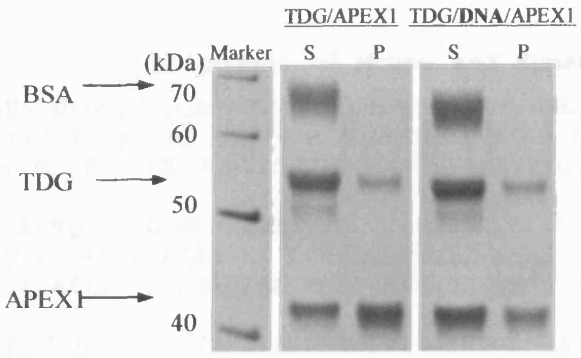



**Figure 62 Pulldown assay of TDG and His-APEX1 in the absence and presence of CpG·T oligonucleotide, highlighting no effect of DNA**

**555 nM TDG and 493 nM His-APEX1 were incubated with nickel-charged resin (1 µl per 15 µl of reaction) for 1h at 4 °C, in the absence (middle panel) or presence (right panel) of 400 nM CpG·T oligonucleotide. S stands for supernatant of the mixture after incubation, and P stands for pellets (i.e. the agarose resin) after the extensive wash.**

Tini *et al.* (2002) reported that mouse TDG with deletion of the N-terminal first 91 amino acids could still interact with APEX1, but that the further deletion of amino acids 1-121 prevented the interaction. They suggested that residues 92-121 of mouse TDG are required for the protein-protein interaction with APEX1 (Tini *et al.*, 2002). This region corresponds to amino acids 82-110 of human TDG (see Figure 63). To test if this region of human TDG interacted with APEX1, a pull-down assay was done incubating His-APEX1 with a chemically synthesized peptide (purchased from Cambridge Research Biochemicals Limited) corresponding to amino acids 82- 110 of human TDG (Figure 64). The peptide had an N-terminal biotin tag and was pulled down using streptavidin beads. There was no non-specific binding of His-APEX to the streptavidin beads (Figure 64).


**Figure 63 Amino acids sequence alignment of Human and Mouse Thymine-DNA Glycosylase**

**The red corresponds to 92-121 amino acids of mouse TDG, which Tini *et al.* (2002) made a point of being responsible for interaction of mouse TDG and APEX1. The pink corresponds to 82-110 amino acids of human TDG which is a peptide sequence used in the pull-down assay (Figure 64).**

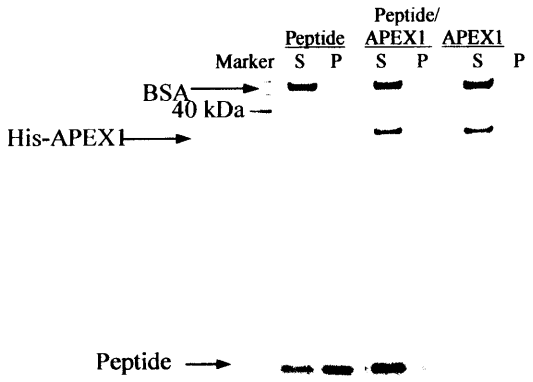



**Figure 64 Pulldown assay of peptide (82-110 amino acids) of TDG with His-APEX1**

**15 μM of biotin-tagged peptide, 493 nM of His-APEX1, and 1 μl of streptavidin beads/15μl of reaction mixture were used. Three reactions (peptide alone with the beads, peptide & His-APEX1 with the beads, and His-APEX1 alone with the beads) were incubated 1h at 4 °C. S stands for supernatant of the mixture after incubation, and P stands for pellets (i.e. the beads) after the extensive wash. Samples were heated in SDS-loading buffer for 10 min and 90 °C before loading onto a SDS-PAGE. Gels were silver-stained.**

The peptide of TDG failed to pull down His-APEX1 (pellet lane of mixture of peptide and His-APEX1 in Figure 64). There are several possible explanations for this. One is that His-APEX1 does not bind to this region of human TDG. Alternatively, the peptide sequence does contain the His-APEX1 binding site but it has to be in the correct three-dimensional structure to be recognized by His-APEX1. If the peptide is not folded correctly (which is quite possible) then APEX1 cannot interact. Another possibility is that His-APEX1 does interact with this region of TDG but requires additional contacts outside of this region to give strong enough binding to be detected by the pull-down assay.

*Summary*

The pulldown assay showed a weak protein-protein interactions between TDG and His-APEX1. In agreement with the results on the effect of oligonucleotide length on the displacement of TDG which showed that APEX1 does not require the DNA as an absolute factor for displacement of TDG from abasic sites, the presence of CpG·T DNA did not significantly increase protein-protein interaction in this pulldown assay. The peptide (82-110) of human TDG was also tested to see if this region was responsible for the interaction of human TDG with His-APEX1, as suggested by the experiments of Tini *et al.* on mouse TDG (Tini *et al.*, 2002). The peptide failed to pull down any APEX1, suggesting that APEX1 might recognize the three-dimensional structure of TDG that requires the rest of the sequence in addition to the peptide, or that this region is not involved in the interaction.

Only a small fraction of TDG was pulled down by His-APEX1 (lane 6 in Figure 61). One possible reason for this weak band of TDG with His-APEX1 could be the very transient nature of the interaction between TDG and His-APEX1. This is supported by the fact that although a bandshift assay showed a complex of TDG and DNA, and another complex of APEX1 and DNA, it failed to detect a ternary complex of TDG, DNA, and APEX1 (Waters *et al.*, 1999). One possible way to detect the ternary complex would be to use an abasic site analogue that is not cleaved by APEX1 (e.g. phosphorothioate) or an APEX1 mutant that cannot cleave abasic sites (e.g. Asp210Ala (Rothwell *et al.*, 2000)). This increases the chance of detecting a ternary complex that forms at the intact abasic site.

## 5.3.2. Isothermal titration calorimetry

To complement the pulldown assay, detection of protein-protein interaction was attempted using isothermal titration calorimetry (ITC; MicroCal$^{TM}$). ITC measures tiny heat difference generated on association of two proteins. The calorimeter contains two cells, a reference cell and a sample cell. The reference cell that only contained buffer and the sample cell that contained TDG were equilibrated at 30 ℃ before adding His-APEX1 stepwise to the sample cell. When APEX1 was added, the heat difference between the reference cell and the sample cell was monitored. In order to carry out one experiment with ITC, at least 1 ml of 60 μM APEX1 and 2 ml of 6 μM TDG were required. Unfortunately, as shown in Section 4.1. and by Tini *et al.*, purification of large quantities of TDG at high concentration exhibited great difficulty since the majority of expressed TDG was easily lost by degradation through purification steps and because TDG precipitates out easily. In this project enough TDG was purified to attempt two ITC experiments. No interaction between TDG and APEX1 was detected (Figure 65) in this limited number of experiments.

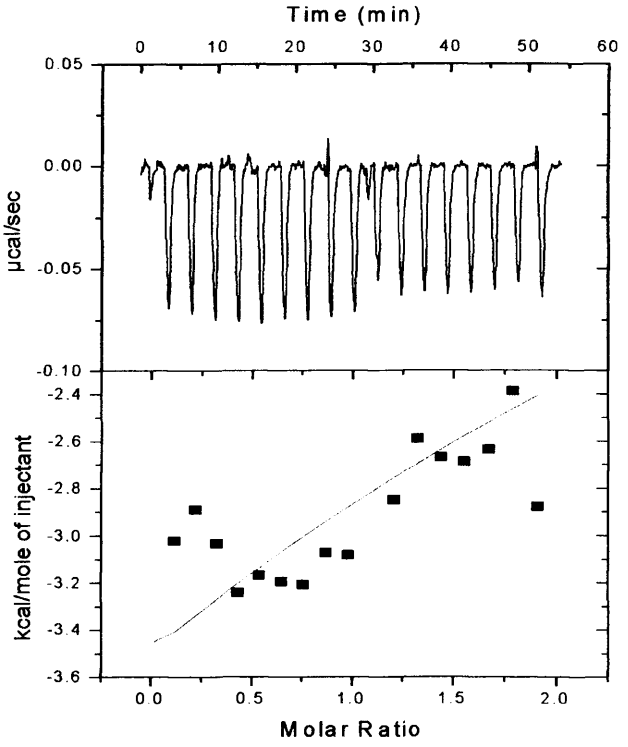



**Figure 65 Isothermal titration calorimetry to detect protein-protein interactions between TDG and His-APEX1 at 30 °C**

**Typical positive result of protein-protein interactions should be a sigmoid shape (in the lower graph) as molar ratio increases. In this experiment, it failed to show the shape.**

However, it is very difficult to analyse this result and conclude that there was no interaction between the two enzymes. To give a positive result, the interaction between two proteins must produce a large enough heat difference to be detectable by ITC. If TDG and APEX1 do interact, but the heat produced on association of the proteins is too small, ITC would not detect the heat difference, giving a negative result. Therefore, failure of the detection could be a machinery

limit, or experimental condition. Further experiments using higher concentration of the proteins to amplify the signal, could be done to see if the negative result found in this project is due to a small heat change on association of TDG with APEX1.

There is a possibility that *in vivo* TDG may need post-translational modification in order to enhance its interaction with APEX1. For example, Hardeland *et al.* (2002) suggested that modification of TDG by ubiquitin-like proteins SUMO-1 and SUMO-3 might enhance the interaction with APEX1. In addition to this sumoylation, acetylation of TDG has also been reported to change the affinity of TDG for APEX1 (Tini *et al.*, 2002). These reports suggest that there might be further regulatory mechanisms that modify TDG (or APEX1) in order to enhance the interaction between the two enzymes in vivo. Lack of the modification in the *in vitro* pull down experiments might lead be the reason for the weak interaction found in this project. *In vivo*, displacement of TDG happens on the DNA strand. Although requirement of DNA for displacement of TDG by APEX1 was not absolute (from the experiments with the different length oligonucleotides), the DNA should also be included in the ITC and pull down experiments. However, because there would be non-specific binding to the DNA and because the transient nature of the interaction, the detection of a specific protein-DNA-protein complex from such a three component experiment would be very difficult. One way to simplify the experiment would be to use TDG cross-linked to DNA since this would eliminate a major part of the ambiguous signal in ITC. Alternatively, DNA containing an un-cleavable abasic site analogue that cannot be processed by TDG may be needed for the pull-down assay and ITC. This should enhance the chance

of detecting the interaction since the un-cleavable DNA generates an equilibrium between TDG, DNA and APEX1, whereas a normal abasic site would be cut by APEX1 and it is probable that TDG would interact weakly with this cut abasic site.

Finally, it is known that other repair proteins are coordinated to protect unstable intermediates in base excision repair; APEX1 interacts with DNA polymerase β (Bennett *et al.*, 1997) and DNA polymerase β interacts with DNA ligase III via XRCC1 (Cappelli *et al.*, 1997; Nash *et al.*, 1997). By showing an interaction between TDG and APEX1, this project completes the picture for base excision repair of G·T mismatches. Thus, the entire repair of G·T mismatches involves the sequential interaction of all of the enzymes in this pathway at the site of repair and this coordination ensures that the unstable repair intermediates are protected. This situation contrasts with the large complexes that comprise multiple proteins at the damaged site seen in nucleotide excision repair and in mismatch repair (Section 2.3.3 and 2.3.4).

# 6. CONCLUSIONS

When 5-methylcytosine spontaneously deaminates, the base changes to thymine, generating a G·T mismatch. TDG was discovered as a glycosylase that excises thymine mismatched with guanine, which then initiates the base excision repair pathway.

In the first part of this project, the kinetic parameters, $k_2$ and $K_d$, of TDG for thymine substrates (CpG·T, and TpG·T) and 3,$N^4$-ethenocytosine substrates (CpG·εC, TpG·εC, GpG·εC, and ApG·εC) were studied. TDG showed a strong preference for thymine in a CpG context, displaying the fastest catalytic constant among the substrates tested. This strong preference supports the observation that TDG has specific contacts with the guanine of the C·G base pair 5' to the mismatch (Scharer *et al.*, 1997). The structure of the *E.coli* homologue of TDG strongly suggests that TDG also contacts the guanine of the G·T mismatch after flipping out the thymine. The extended contact of TDG with the 5'-adjacent C·G base pair has probably developed through evolution to specifically excise thymine generated by deamination of 5-methylcytosine at CpG sites, to maintain the integrity of the genome. But what is the explanation for the relatively slow excision of thymine by TDG? Unlike uracil (the substrate of UDG), and many other base lesions that are not normally found in DNA, the thymine substrate of TDG is a normal DNA base. The vast majority of thymine in the genome is paired with adenine. Therefore, TDG has to discriminate between rare thymines in G·T mismatches from the far more common thymines that are in A·T base pairs. *In vitro*, TDG excises thymine from A·T base pairs at an extremely slow rate (Waters and Swann, 1998), partly because the glycosylase makes contact to the mismatched guanine of a G·T mismatch, but probably also because it is harder for TDG to flip out the thymine from the more stable A·T base pair. However, this level of discrimination may not be enough *in vivo* and so to avoid removing thymine unnecessarily from A·T base pairs, TDG may have evolved to have a slow thymine excision rate.

*In vitro*, TDG can excise a broad range of modified bases (Section 2.4.3.). It has been claimed that ethenocytosine is the real substrate of TDG (Hang *et al.*, 1998; Saparbaev and Laval, 1998), since these studies seemed to find that TDG reacts faster on ethenocytosine than thymine. However, these studies were flawed

because they did not look at the effect of neighbouring DNA sequence on ethenocytosine excision nor did they allow for product inhibition. By doing a very careful kinetic analysis of single turnover base excision by TDG, the results of this project show that the excision of ethenocytosine is very dependent upon the neighbouring base pair. In terms of the specificity constant, $k_2/K_d$, the order of preference is: CpG·εC >> TpG·εC ≈ GpG·εC > ApG·εC. The specificity constant for the best ethenocytosine substrate, CpG·εC, was more than 1600-fold greater than the worst substrate ApG·εC. This means that ethenocytosine in a CpG site would be repaired most efficiently by TDG, but that ethenocytosine in an ApG site would be repaired at a very poor rate. Consequently, if TDG were the main enzyme responsible for repairing ethenocytosine in the cell, ethenocytosines formed at ApG sites would go un-repaired. This would not be a physiologically satisfactory situation for cells since there is no evidence that ethenocytosine is formed exclusively at CpG sites. Thus, it is very unlikely that TDG is the major *in vivo* activity responsible for ethenocytosine excision. However, there is the anomaly that ethenocytosine seems a better substrate than thymine on the basis of specificity constants. The large specificity constant for CpG·εC comes from its $K_d$ being nearly 800-fold smaller than the $K_d$ of CpG·T even though the catalytic rate constant of ethenocytosine is approximately six times slower than that of thymine. From the inhibition studies done in this project, together with the fact that melting temperature and structural studies have shown that a G·εC base pair is weaker than a G·T mismatch, it can be concluded that most of the 800-fold difference in $K_d$ comes from it being much easier for TDG to flip out ethenocytosine from G·εC base pairs than to flip out thymine from G·T mismatches. It is generally believed that evolution directs enzymes to reduce the activation energy of the reaction they catalyse, to be more efficient. A consequence of this is that binding energy is utilized to increase the catalytic rate constant ($k_2$) rather than to enhance binding to the substrate (and lower $K_d$). Thus the apparent fast reaction of ethenocytosine due to smaller $K_d$ is probably accidental to TDG since the catalytic rate constant is smaller (≈ larger activation energy) than that of thymine. This again suggests that it is unlikely that ethenocytosine is the primary in vivo substrate of TDG.

The presence of APEX1 increases the turnover rate of TDG by displacing the glycosylase from its abasic site product, but we do not know how APEX1 achieves this displacement The involvement of DNA and protein-protein interactions in the displacement of TDG by APEX1 was examined in the second part of this project APEX1 displacement of TDG was reduced when oligonucleotide either side of the TDG-DNA complex was not available for APEX1, but none of oligonucleotides tested completely abolished it, including the shortest oligonucleotide that would be entirely covered by TDG. Thus, free DNA base pairs certainly help to load APEX1 onto the complex, but they are not an absolute requirement of APEX1 for the displacement of TDG. A protein-protein interaction between human TDG and APEX1 was found using a pulldown assay but was not detectable in preliminary isothermal calorimetry experiments. Therefore, TDG probably does interact directly with APEX1 but this interaction is weak or transient.

From the results the following possible mechanism for the displacement of TDG by APEX1 is proposed. APEX1 interacts with the abasic-site-bound TDG, which induces a conformational change in TDG. The conformational change loosens the contact between TDG and DNA including flipped out abasic site, widowed guanine and the C·G base pair 5' to the mismatch, causing the glycosylase to dissociate. If APEX1 only interacts with the DNA bound TDG, this interaction would be transient and would thus explain the weak pulldown band. APEX1 probably has this interaction with TDG to protect the abasic site from exposure to the environment, which would occur if TDG released the abasic DNA spontaneously. From the results of displacement experiment using different length oligonucleotides, it is almost impossible for APEX1 to distort DNA structure before release of TDG from the abasic site. Therefore, the proposed mechanism of APEX1 distorting the DNA to displace TDG cannot be true. After release of TDG, the abasic site taken over by APEX1 undergoes into the distorted form as seen in the crystal structure of APEX1 bound to DNA (Gorman *et al.*, 1997).

# 7. ABBREVIATIONS

| | |
|---|---|
| A | Adenine |
| AP endonuclease | Apurinic/apyrimidinic endonuclease |
| APEX1 | Human apurinic/apyrimidinic endonuclease 1 |
| BER | Base excision repair |
| bp | Base pair |
| BSA | Bovine serum albumin |
| C | Cytosine |
| cDNA | Complementary DNA |
| CHO | Chinese hamster ovary |
| DNA | Deoxynucleic acid |
| DTT | Dithiothreitol |
| εC | 3,$N^4$-Ethenocytosine |
| *E.coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraaceticacid |
| FPLC | Fast protein liquid chromatography |
| G | Guanine |
| HEPES | N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonicacid |
| hNth1 | Human thymine glycol DNA glycosylase-AP lyase |
| hOGG1 | Human 8-oxo-7,8-dihydroguanine DNA glycosylase |
| HPLC | High performance liquid chromatography |
| hUDG | Human uracil DNA glycosylase |
| Ig | Immunoglobulin |
| ITC | Isothermal titration calorimetry |
| LB medium | Luria-Bertani medium |
| $^{5\text{-me}}$C | 5-Methylcytosine |

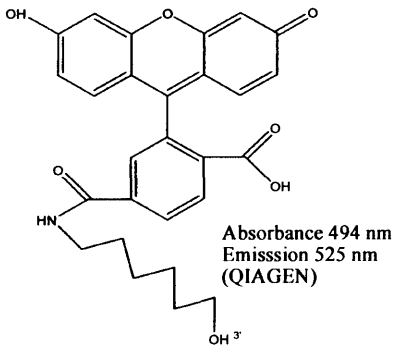
| | |
|---|---|
| MMR | Mismatch repair |
| MUG | Mismatch-specific uracil-DNA glycosylase |
| mRNA | Messenger RNA |
| NER | Nucleotide excision repair |
| NMR | Nuclear magnetic resonance |
| PCNA | Proliferating cell nuclear antigen |
| RAR | Retinoid acid receptor |
| RXR | Retinoid X receptor |
| SDS | Sodium dodecyl sulfate |
| T | Thymine |
| TCR | Transcription-coupled repair |
| TAE | Tris-acetate electrophoresis |
| TBE | Tris-borate electrophoresis |
| TDG | Thymine-DNA Glycosylase |
| UDG | Uracil-DNA glycosylase |

# 8. APPENDICES

# 8.1. Structure of 6-FAM

6-FAM was incorporated at 5'-end of the bottom strand of oligonucleotides. The fluorescent labelled oligonucleotides were purchased from Qiagen.

**6-FAM**




Absorbance 494 nm
Emisssion 525 nm
(QIAGEN)

# 8.2. Sequences of Human Thymine DNA Glycosylase and Apurinic/apyrimidinic Endonuclease 1

## 8.2.1. The cDNA and amino acids sequences of human thymine DNA glycosylase

Nucleotide sequence:

```
0       atggaagcgg agaacgcggg cagctattcc cttcagcaag ctcaagcttt
51      ttatacgttt ccatttcaac aactgatggc tgaagctcct aatatggcag
101     ttgtgaatga acagcaaatg ccagaagaag ttccagcccc agctcctgct
151     caggaaccag tgcaagaggc tccaaaagga agaaaaagaa aacccagaac
201     aacagaacca aaacaaccag tggaacccaa aaaacctgtt gagtcaaaaa
251     aatctggcaa gtctgcaaaa tcaaaagaaa aacaagaaaa aattacagac
301     acatttaaag taaaaagaaa agtagaccgt tttaatggtg tttcagaagc
351     tgaacttctg accaagactc tccccgatat tttgaccttc aatctggaca
401     ttgtcattat tggcataaac ccgggactaa tggctgctta caaagggcat
451     cattaccctg gacctggaaa ccatttttgg aagtgtttgt ttatgtcagg
501     gctcagtgag gtccagctga accatatgga tgatcacact ctaccaggga
551     agtatggtat tggatttacc aacatggtgg aaaggaccac gcccggcagc
601     aaagatctct ccagtaaaga atttcgtgaa ggaggacgta ttctagtaca
651     gaaattacag aaatatcagc cacgaatagc agtgtttaat ggaaaatgta
701     tttatgaaat ttttagtaaa gaagtttttg gagtaaaggt taagaacttg
751     gaatttgggc ttcagcccca taagattcca gacacagaaa ctctctgcta
801     tggtatgcca tcatccagtg caagatgtgc tcagtttcct cgagcccaag
851     acaaagttca ttactacata aaactgaagg acttaagaga tcagttgaaa
901     ggcattgaac gaaatatgga cgttcaagag gtgcaatata catttgacct
951     acagcttgcc caagaggatg caaagaagat ggctgttaag gaagaaaaat
1001    atgatccagg ttatgaggca gcatatggtg gtgcttacgg agaaaatcca
1051    tgcagcagtg aaccttgtgg cttctcttca aatgggctaa ttgagagcgt
1101    ggagttaaga ggagaatcag ctttcagtgg cattcctaat gggcagtgga
1151    tgacccagtc atttacagac caaattcctt cctttagtaa tcactgtgga
1201    acacaagaac aggaagaaga aagccatgct taa
```

Amino acids sequence:

MEAENAGSYSLQQAQAFYTFPFQQLMAEAPNMAVVNEQQMPEEVPAPAPAQEPVQ
EAPKGRKRKPRTTEPKQPVEPKKPVESKKSGKSAKSKEKQEKITDTFKVKRKVDR
FNGVSEAELLTKTLPDILTFNLDIVIIGINPGLMAAYKGHHYPGPGNHFWKCLFM
SGLSEVQLNHMDDHTLPGKYGIGFTNMVERTTPGSKDLSSKEFREGGRILVQKLQ
KYQPRIAVFNGKCIYEIFSKEVFGVKVKNLEFGLQPHKIPDTETLCYGMPSSSAR
CAQFPRAQDKVHYYIKLKDLRDQLKGIERNMDVQEVQYTFDLQLAQEDAKKMAVK
EEKYDPGYEAAYGGAYGENPCSSEPCGFSSNGLIESVELRGESAFSGIPNGQWMT
QSFTDQIPSFSNHCGTQEQEEESHA

## 8.2.2. The cDNA and amino acids sequences of human apurinic/apyrimidinic endonuclease 1

**Nucleotide sequence:**

```
0     atgccgaagc gtgggaaaaa gggagcggtg gcggaagacg gggatgagct
51    caggacagag ccagaggcca agaagagtaa gacggccgca aagaaaaatg
101   acaaagaggc agcaggagag ggcccagccc tgtatgagga ccccccagat
151   cagaaaacct cacccagtgg caaacctgcc acactcaaga tctgctcttg
201   gaatgtggat gggcttcgag cctggattaa gaagaaagga ttagattggg
251   taaaggaaga agccccagat atactgtgcc ttcaagagac caaatgttca
301   gagaacaaac taccagctga acttcaggag ctgcctggac tctctcatca
351   atactggtca gctccttcgg acaaggaagg gtacagtggc gtgggcctgc
401   tttcccgcca gtgcccactc aaagtttctt acggcatagg cgatgaggag
451   catgatcagg aaggccgggt gattgtggct gaatttgact cgtttgtgct
501   ggtaacagca tatgtaccta atgcaggccg aggtctggta cgactggagt
551   accggcagcg ctgggatgaa gcctttcgca agttcctgaa gggcctggct
401   tcccgaaagc cccttgtgct gtgtggagac ctcaatgtgg cacatgaaga
451   aattgacctt cgcaacccca aggggaacaa aaagaatgct ggcttcacgc
501   cacaagagcg ccaaggcttc ggggaattac tgcaggctgt gccactggct
551   gacagcttta ggcacctcta ccccaacaca ccctatgcct acacctttg
601   gacttatatg atgaatgctc gatccaagaa tgttggttgg cgccttgatt
651   acttttgtt gtcccactct ctgttacctg cattgtgtga cagcaagatc
701   cgttccaagg ccctcggcag tgatcactgt cctatcaccc tatacctagc
751   actgtga
```

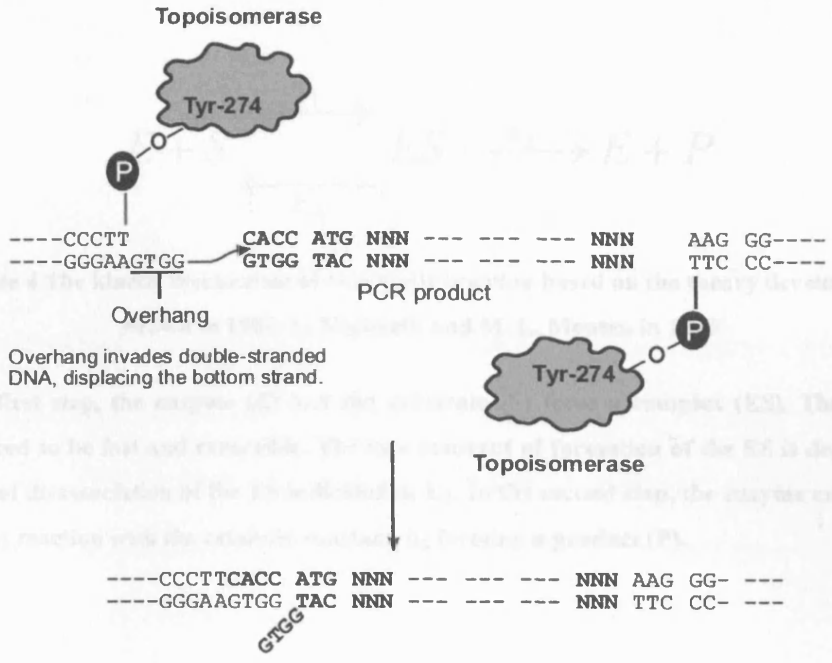
Amino acids sequence:

MPKRGKKGAVAEDGDELRTEPEAKKSKTAAKKNDKEAAGEGPALYEDPPDQKTSP
SGKPATLKICSWNVDGLRAWIKKKGLDWVKEEAPDILCLQETKCSENKLPAELQE
LPGLSHQYWSAPSDKEGYSGVGLLSRQCPLKVSYGIGDEEHDQEGRVIVAEFDSF
VLVTAYVPNAGRGLVRLEYRQRWDEAFRKFLKGLASRKPLVLCGDLNVAHEEIDL
RNPKGNKKNAGFTPQERQGFGELLQAVPLADSFRHLYPNTPYAYTFWTYMMNARS
KNVGWRLDYFLLSHSLLPALCDSKIRSKALGSDHCPITLYLAL

# 8.3. Directional TOPO Cloning System (Invitrogen)

In order to attach histidine tag at the N-terminal of APEX1, this directional TOPO cloning system was used.

A 3'-overhang (GTGG) in the cloning vector invades the 5'-end of a PCR product, and stabilizes the PCR product in the correct orientation.



# 8.4. Kinetics of chemical reactions

The kinetic mechanism of chemical reactions was developed by Brown (1902), then Henri (1913), Michaelis and Menten (1913) (Copeland, 2000; Palmer, 1995). They rationalized that enzyme molecule binds to substrate molecule to form an enzyme-substrate complex. Some of the complex proceeds to chemical reaction, and forming a product molecule. However, some of them fall apart into enzyme and substrate molecules back again (See Scheme 4).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

**Scheme 4 The kinetic mechanism of enzymatic reaction based on the theory developed by Brown in 1902, L. Michaelis and M. L. Menten in 1913**

**At the first step, the enzyme (E) and the substrate (S) form a complex (ES). This step is considered to be fast and reversible. The rate constant of formation of the ES is depicted as $k_1$, and of disassociation of the ES indicated as $k_{-1}$. In the second step, the enzyme catalyses a chemical reaction with the catalytic constant, $k_2$ forming a product (P).**

According to Scheme 4, formation of product depends on the amount of enzyme-substrate complex. At low concentration of substrate, the enzyme molecules are far from "saturated" even though the concentration of substrate is excess over the enzyme. Until all enzyme molecules are saturated to form the enzyme-substrate complex, formation of product is proportional to concentration of substrate at constant enzyme concentration. From the certain level of substrate concentration, *i.e.* high enough to saturate all enzyme molecules to form ES complex, the rate of reaction depends on the chemical reaction to produce product, not the substrate

concentration. That is, further increase of substrate concentration does not make any difference on the rate of reaction in that condition. In other words, it reaches the maximum initial velocity ($V_{max}$) at the enzyme concentration;

$$V_{max} = k_2 [ES]$$

Since the enzyme-substrate concentration, [ES] is equal to the total enzyme concentration, $[E_0]$, the equation can be rewritten;

$$V_{max} = k_2[E_0]$$

Henri (1913), Michaelis and Menten (1913) developed mathematical formula to express the kinetic mechanism (Copeland, 2000). They assume that formation of ES complex and breakdown of the complex to E and S are rapid and in an equilibrium (See Scheme 4). The equilibrium is instantly established, and formation of product from the ES complex is too slow to disturb the equilibrium. When you give an equation based on the assumption,

$$k_1[E][S] = k_{-1}[ES]$$

The constants can be separated to express another constant, $K_d$, the dissociation constant of the ES complex.

$$\frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1} = K_d$$

The concentration of free enzyme, [E], should be the amount that subtracts the concentration of bound enzyme, [ES], from total concentration of enzyme, [$E_0$];

$$[E] = [E_0] - [ES]$$

$$\frac{([E_0]-[ES])[S]}{[ES]} = K_d$$

$$K_d[ES] = ([E_0] - [ES])[S]$$

$$K_d[ES] = [E_0][S] - [ES][S]$$

$$[ES][S] + K_d[ES] = [E_0][S]$$

$$[ES]([S] + K_d) = [E_0][S]$$

$$[ES] = \frac{[E_0][S]}{[S]+K_d}$$

Since the overall rate of reaction is expressed as $v_0 = k_2$ [ES] under a saturation of enzyme;

$$v_0 = \frac{k_2[E_0][S]}{[S]+K_d}$$

and, $V_{max} = k_2[E_0]$, the above equation is rewritten as;

$$v_0 = \frac{V_{max}[S]}{[S]+K_d}$$

This equation formed by Michaelis and Menten was further developed by Briggs and Haldane in 1925 (Copeland, 2000; Palmer, 1995). They introduced more generalized assumption on the kinetic mechanism (Scheme 4). They assumed that the rate of formation of the ES complex is equivalent to the rate of breakdown to reactants and products rather than the equilibrium between reactants and ES complex, which requires $k_2 << k_{-1}$. In this new assumption, it is said that the ES complex exists at steady state concentration. Under this steady-state assumption, the formula is changed to;

$$k_1[E][S] = k_{-1}[ES] + k_2[ES]$$

$$k_1[E][S] = [ES](k_{-1} + k_2)$$

then, separating the constants to obtain another constant, $K_m$, called Michaelis constant, however, when $k_{-1} >> k_2$, $K_m$ becomes equal to $K_d$;

$$\frac{[E][S]}{[ES]} = \frac{k_{-1} + k_2}{k_1} = K_m$$

since the free [E] is the amount that subtracts the [ES] from the total enzyme, $[E_0]$;

$$\frac{([E_0] - [ES])[S]}{[ES]} = K_m$$

$$[ES] = \frac{[E_0][S]}{[S] + K_m}$$

then, the $v_0 = k_2[ES]$ is applied to the above equation;

$$v_0 = \frac{k_2[E_0][S]}{[S]+K_m}$$

also, $V_{max} = k_2[E_0]$ is put on the equation;

$$v_0 = \frac{V_{max}[S]}{[S]+K_m}$$

This equation of enzyme kinetics is called the Michaelis-Menten equation.

$K_m$ is interpreted as a measure of overall breakdown of the ES complex to reactants and products. The unit of $K_m$ is molarity (same unit of substrate) since the value of $K_m$ comes from first order rate constants ($k_{-1}$ and $k_2$ have units of reciprocal time) over second order rate constant ($k_1$ has unit of reciprocal molarity and reciprocal time). If we carry out the kinetic experiment with concentration of substrate that equals $K_m$ value, we will see that the initial velocity will reach at 1/2 of $V_{max}$. The equation is below;

$$v_0 = \frac{V_{max}[S]}{[S]+[S]} = \frac{1}{2}V_{max}$$

In other words, $K_m$ is the substrate concentration that attains half of the maximum velocity, and the substrate concentration at which half of the enzyme molecules in the reaction mixture are saturated with substrates in the steady state.

$K_d$ is the ratio of the reverse over forward reaction rates, and is considered as a dissociation constant of the ES complex. That is, $K_d$ is a gauge of affinity of the

enzyme for the substrate. It is useful quantity to compare between different substrates with the same enzyme. The unit of $K_d$ is also molarity.

The catalytic constant (or turnover number), $k_{cat}$ represents the rate of chemical reaction that may involve multiple steps including a release of product (see Scheme 5). For the multiple-complex reaction, $k_{cat}$ is a function of $k_2$ and $k_3$. The $k_{cat}$ can be obtained by dividing the value of $V_{max}$ by the concentration of enzyme since $V_{max} = k_{cat}[E_0]$. The unit of $k_{cat}$ is a reciprocal time.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} ES' \xrightarrow{k_3} E + P$$

**Scheme 5 The kinetic mechanism that involves multiple catalytic steps**

**The ES complex changes to the transient complex ES' with the constant, $k_2$, and then the ES' becomes E and P with $k_3$.**

In case of TDG reaction, $k_{cat}$ cannot be obtained since TDG is not released from the product (see Scheme 6). That is, TDG shows a single turnover reaction.

$$TDG + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} TDG - S \xrightarrow{k_2} TDG - P$$

**Scheme 6 Schematic mechanism of excision of a base by thymine DNA glycosylase**

**S stands for substrate, and P stands for product. One molecule of TDG binds to one molecule of substrate, forming a complex TDG-S with the rate constant, $k_1$. Formation of the complex is reversed with the rate constant, $k_{-1}$. Some of the complex undergoes chemical reaction, cleavage of glycosydic bond of mismatched base, with the rate constant, $k_2$. After the mismatched base is excised, TDG remains at the abasic site. Thus, there is no further reaction of the enzyme unless a releasing factor is present.**

Formation and breakdown of TDG-S complex is rapid, and a limiting step of the reaction is the chemical cleavage of the mismatched base, i.e. $k_2$ in Scheme 6. The dissociation constant, $K_d$ ( = $k_{-1}$ / $k_1$) and the catalytic constant, $k_2$ were calculated by measuring amount of product formed during the initial period of the reaction but in a steady state.

The specificity constant (or catalytic efficiency), $k_{cat}/K_d$ (or $K_m$) is used to compare the efficiencies of different enzymes, and different substrates for one enzyme. A high value of $k_{cat}/K_d$ manes that it reacts quickly as soon as the ES complex is formed, then the limiting aspect of the reaction tends to be a matter of collision between enzyme molecule and substrate molecule. A low value of the constant becomes a close situation to the equilibrium assumption, i.e. $k_{-1} >> k_2$ (Gutfreund, 1995; Palmer, 1995).

In this project, $K_d$, and $k_2$ derived from the model (Scheme 6) were obtained for thymine and ethenocytosine substrates under a single-turnover reaction.

# 9. ACKNOWLEDGEMENTS

To my friend Kazumi in memoriam

I would like to thank Dr. Timothy R. Waters for his guidance and encouragement throughout the project. I appreciate Nicola Brookman-Amissah, Ray Mace, Peter Swann and all my friends for their assistance. I also appreciate people in Dr. John Ladbury's group for their support in ITC experiment.

Above all, I would like to thank my parents, Kaichi and Yoko, and the rest of my family since I would not complete the project without their support.

# 10. REFERENCES

Abu, M., and Waters, T. R. (2003). The Main Role of Human Thymine-DNA Glycosylase Is Removal of Thymine Produced by Deamination of 5-Methylcytosine and Not Removal of Ethenocytosine. J. Biol. Chem. 278, 8739-8744.

Alekseyev, Y. O., and Romano, L. J. (2000). In vitro replication of primer-templates containing benzo[a]pyrene adducts by exonuclease-deficient Escherichia coli DNA polymerase I (Klenow fragment). Biochemistry 39, 10431-10438.

Allawi, H. T., and SantaLucia, J., Jr. (1998). NMR solution structure of a DNA dodecamer containing single G·T mismatches. Nucleic Acids Res. 26, 4925-4934.

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., and Wyllie, A. (1999). Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. Oncogene 18, 8044-8047.

Balylin, S. B., and Herman, J. G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet. 16, 168-174.

Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. Cell 92, 117-129.

Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999). Crystal structure of a thwarted mismatch glycosylase DNA repair complex. EMBO J. 18, 6599-6609.

Bartsch, H., and Nair, J. (2000). New DNA-based biomarkers for oxidative stress and cancer chemoprevention studies. Eur. J. Cancer 36, 1229-1234.

Barzilay, G., and Hickson, I. D. (1995). Structure and function of apurinic/apyrimidinic endonucleases. BioEssays 17, 713-719.

Begley, T. J., and Samson, L. D. (2003). AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. Trends Biochem. Sci. 28, 2-5.

Bellacosa, A. (2001). Role of MED1 (MBD4) Gene in DNA repair and human cancer. J. Cell. Physiol. 187, 137-144.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc. Nat. Acad. Sci. U. S. A. 96, 3969-3974.

Bellamy, S. R. W., and Baldwin, G. S. (2001). A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1. Nucleic Acids Res. 29, 3857-3863.

Bennett, R. A. O., Wilson, D. M., III, Wong, D., and Demple, B. (1997). Interaction of human apurinic endonuclease and DNA polymerase beta in the base excision repair pathway. Proc. Natl. Acad. Sci. U. S. A. 94, 7166-7169.

Bestor, T. H. (1998). Gene silencing. Methylation meets acetylation [news; comment]. Nature 393, 311-312.

Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: develpment of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 58, 5248-5257.

Bouziane, M., Miao, F., Bates, S. E., Somsouk, L., Sang, B. C., Denissenko, M., and O'Connor, T. R. (2000). Promoter structure and cell cycle dependent expression of the human methylpurine-DNA glycosylase gene. Mut. Res. 461, 15-29.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Annu. Biochem. 72, 248-254.

Buermeyer, A. B., Deschenes, S. M., Baker, S. M., and Liskay, R. M. (1999). MAMMALIAN DNA MISMATCH REPAIR. Annu. Rev. Genet. 33, 533-564.

Caldecott, K. W., McKeown, C. K., Tucker, J. D., Ljungquist, S., and Thompson, L. H. (1994). An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. Mol. Cell Biol. 14, 68-76.

Cappelli, E., Taylor, R., Cevasco, M., Abbondandolo, A., Caldecott, K., and Frosina, G. (1997). Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. J. Biol. Chem. 272, 23970-23975.

Chen, D., Lucey, M. J., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Losson, R., Buluwela, L., Coombes, R. C., Chambon, P., Schar, P., and Ali, S. (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. J. Biol. Chem. 278, 38586-38592.

Chen, D. S., Herman, T., and Demple, B. (1991). Two distinct human DNA diesterases that hydrolyze 3'-blocking deoxyribose fragments from oxidized DNA. Nucleic Acids Res. 19, 5907-5914.

Chevray, P. M., and Nathans, D. (1992). Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. Proc. Natl. Acad. Sci. U. S. A. 89, 5789-5793.

Chrivia, J. C., Kwok, R. P., Lamb, N., Hagiwara, M., Montminy, M. R., and Goodman, R. H. (1993). Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature 365, 855-859.

Chung, F.-L., Chen, H.-J. C., and Nath, R. G. (1996). Lipid peroxidation as a potential endogenous source for the formation of exocyclic DNA adducts. Carcinogenesis 17, 2105-2111.

Cooper, P. K., Nouspikel, T., Clarkson, S. G., and Leadon, S. (1997). Defective Transcription-Coupled Repair of Oxidative Base Damage in Cockayne Syndrome Patients from XP Group G. Science 275, 990-993.

Copeland, R. A. (2000). ENZYMES A Practical Introduction to Structure, Mechanism, and Data Analysis, Second Edition Edition: WILEY-VCH.

Cross, S. H., and Bird, A. P. (1995). CpG islands and genes. Curr. Opin. Genet. Dev. 5, 309-314.

Cullinan, D., Johnson, F., Grollman, A. P., Eisenberg, M., and de los Santos, C. (1997). Solution structure of a DNA duplex containing the exocyclic lesion 3,N4-etheno-2'-deoxycytidine opposite 2'-deoxyguanosine. Biochemistry 36, 11933-11943.

Delort, A. M., Duplaa, A. M., Molko, D., Teoule, R., Leblanc, J. P., and Laval, J. (1985). Excision of uracil residues in DNA: mechanism of action of Escherichia coli and Micrococcus luteus uracil-DNA glycosylases. Nucleic Acids Res. 13, 319-335.

Denissenko, M. F., Chen, J. X., Tang, M.-S., and Pfeifer, G. P. (1997). Cytosine methylation determines hot spots of DNA damage in the human P53 gene. Proc. Natl. Acad. Sci. U. S. A. 94, 3893-3898.

Devlin, T. M. (1992). TEXTBOOK OF BIOCHEMISTRY with Clinical Correlations, Third Edition Edition: WILEY-LISS.

Dianova, II, Bohr, V. A., and Dianov, G. L. (2001). Interaction of human AP endonuclease 1 with flap endonuclease 1 and proliferating cell nuclear antigen involved in long-patch base excision repair. Biochemistry 40, 12639-12644.

Dietmaier, W., Wallinger, S., Bocker, T., Kullmann, F., Fishel, R., and Ruschoff, J. (1997). Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. Cancer Res. 57, 4749-4756

Dimitriadis, E. K., Prasad, R., Vaske, M. K., Chen, L., Tomkisnson, A. E., Lewis, M. S., and Wilson, S. H. (1998). Thermodynamics of Human DNA Ligase I Trimerization and Association with DNA Polymerase b. J. Biol. Chem. 273, 20540-20550.

Dinner, A. R., Blackburn, G. M., and Karplus, M. (2001). Uracil-DNA glycosylase acts by substrate autocatalysis. Nature 413, 752-755.

Dizdaroglu, M. (2003). Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. Mut. Res. 531, 109-126.

Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G., and Krokan, H. E. (1996). Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. Nucleic Acids Res. 24, 418-422.

Eckner, R., Ewen, M. E., Newsome, D., Gerdes, M., DeCaprio, J. A., Lawrence, J. B., and Livingston, D. M. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. 8, 869-884.

Ehrlich, M., Gama, S. M., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 10, 2709-2721.

Ehrlich, M., Norris, K. F., Wang, R. Y.-H., Kuo, K. C., and Gehrke, C. W. (1986). DNA cytosine methylation and heat-induced deamination. Biosciences Rep. 6, 387-393.

Ehrlich, M., Zhang, X. Y., and Inamdar, N. M. (1990). Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. Mut. Res. 238, 277-286.

Evans, A. R., Limp-Foster, M., and Kelley, M. R. (2000). Going APE over ref-1. Mut. Res. 461, 83-108.

Fersht, A. (1985). Enzyme structure and mechanism, 2nd Edition (New York: W. H. Freeman and Company.

Friedberg, E. C. (1985). DNA Repair: W.H.FREEMAN AND COMPANY.

Friedberg, E. C. (1996). Relationships between DNA repair and transcription. Annu. Rev. Biochem. 65, 15-42.

Friedberg, E. C., Walker, G., and Siede, W. (1995). DNA repair and mutagenesis: ASM Press, Washington.

Frosina, G. (2000). Review Article Overexpression of enzymes that repair endogenous damage to DNA. Eur. J. Biochem. 267, 2135-2149.

Gallinari, P., and Jiricny, J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. Nature 383, 735-738.

Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1998). The impact of an exocyclic cytosine adduct on DNA duplex properties: significant thermodynamic consequences despite modest lesion-induced structural alterations. Biochemistry 37, 12507-12512.

Ghissassi, F. E., Barbin, A., Nair, J., and Bartsch, H. (1995). Formation of 1,N6-Ethenoadenine and 3,N4-Ethenocytosine by Lipid Peroxidation Products and Nucleic Acid Bases. Chem. Res. Toxicol. 8, 278-283.

Gorman, M. A., Morera, S., Rothwell, D. G., de La Fortelle, E., Mol, C. D., Tainer, J. A., Hickson, I. D., and Freemont, P. S. (1997). The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. EMBO J. 16, 6548-6558.

Greenblatt, M. S., Bennett, W. P., Hollstein, M., and Harris, C. C. (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenes. Cancer Res. 54, 4855-4878.

Griffin, S., Branch, P., Xu, Y.-Z., and Karran, P. (1994). DNA mismatch binding and incision at modified guanine bases by extracts of mammalian cells: implications for tolerance to DNA methylation damage. Biochemistry 33, 4787-4793.

Griffin, S., and Karran, P. (1993). Incision at DNA G.T mispairs by extracts of mammalian cells occurs preferentially at cytosine methylation sites and is not targeted by a separate G.T binding reaction. Biochemistry 32, 13032-13039.

Gu, L., Hong, Y., McCulloch, S., Watanabe, H., and Li, G. M. (1998). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. Nucleic Acids Res. 26, 1173-1178.

Gutfreund, H. (1995). Kinetics For The Life Sciences Receptors, transmitters and catalysis: Cambridge University Press.

Hainaut, P., Soussi, T., Shomer, B., Hollstein, M., Greenblatt, M., Hovig, E., Harris, C. C., and Montesano, R. (1997). Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. Nucl Acids Res. 15, 151-157.

Hang, B., Downing, G., Guliaev, A. B., and Singer, B. (2002). Novel Activity of Escherichia coli Mismatch Uracil-DNA Glycosylase (Mug) Excising 8-(Hydroxymethyl)-3,N4-ethenocytosine, a Potential Product Resulting from Glycidaldehyde Reaction. Biochemistry 41, 2158-2165.

Hang, B., Medina, M., Fraenkel-Conrat, H., and Singer, B. (1998). A 55-kDa protein isolated from human cells shows DNA glycosylase activity toward 3,N4-ethenocytosine and the G/T mismatch. Proc. Natl. Acad. Sci. U. S. A. 95, 13561-13566.

Härd, T., and Lundback, T. (1996). Thermodynamics of sequence-specific protein-DNA interactions. Biophys. Chem. 62, 121-139.

Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., and Schar, P. (2001). Thymine DNA glycosylase. Prog. Nucleic Acid Res. Mol. Biol. 68, 235-253.

Hardeland, U., Steinacher, R., Jiricny, J., and Schar, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. EMBO J. 21, 1456-64.

Haug, T., Skorpen, F., Aas, P. A., Malm, V., Skjelbred, C., and Krokan, H. E. (1998). Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. Nucleic Acids Res. 26, 1449-1457.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Molec. and Cellular Biol. 18, 6538-6547.

Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 401, 301-304.

Hill, J. W., Hazra, T. K., Izumi, T., and Mitra, S. (2001). Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. Nucleic Acids Res. 29, 430-438.

Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P., and Tainer, J. A. (1999). Structure of the DNA repair enzyme Endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. Cell 98, 397-408.

Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. (1987). The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. J. Biol. Chem. 262, 9962-9970.

Jiricny, J. (1998). Replication errors: cha(lle)nging the genome. EMBO J. 17, 6427-6436.

Jones, P. A., and Laird, P. W. (1999). Cancer epigenetics comes of age. Nature Genet. 21, 163-167.

Jones, P. A., Rideout, W. M. d., Shen, J. C., Spruck, C. H., and Tsai, Y. C. (1992). Methylation, mutation and cancer. BioEssays 14, 33-36.

Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., and Slupphaug, G. (2002). hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. J. Biol. Chem. 277, 39926-39936.

Kelley, M. R., Cheng, L., Foster, R., Tritt, R., Jiang, J., Broshears, J., and Koch, M. (2001). Elevated and altered expression of the multifunctional DNA base excision repair and redox enzyme Ape1/ref-1 in prostate cancer. Clin. Cancer Res. 7, 824-830.

Kimura, S., Hara, Y., Pineau, T., Fernandez-Salguero, P., Fox, C. H., Ward, J. M., and Gonzalez, F. J. (1996). The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev. 10, 60-69.

Klungland, A., Hoss, M., Gunz, D., Constantinou, A., Clarkson, S. G., Doetsch, P. W., Bolton, P. H., Wood, R. D., and Lindahl, T. (1999). Base excision repair of oxidative DNA damage activated by XPG protein. Mol. Cell 3, 33-42.

Krokan, H. E., Drablos, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. Oncogene 21, 8935-8948.

Krokan, H. E., Standal, R., and Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. Biochem. J. 325, 1-16.

Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E., and Lindahl, T. (1996). Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. EMBO J. 15, 6662-6670.

Kunkel, T. A., and Wilson, S. H. (1996). DNA repair. Push and pull of base flipping. Nature 384, 25-26.

Lahm, A., and Suck, D. (1991). DNase I-induced DNA conformation. 2A structure of a DNaseI-octamer complex. J. Mol. Biol. 222, 645-667.

Modrich, P., and Lahue, M. R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. 65, 101-133.

Lehmann, A. R. (1995). Nucleotide excision-repair and the link with transcription. Trends Biochem. Sci. 20, 402-405.

Lewin, B. (1997). GENES VI: Oxford University Press.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69, 905-914.

Lindahl, T. (1974). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proc. Natl. Acad. Sci. U. S. A. 71, 3649-3653

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature 362, 709-715.

Lindahl, T. (2000). Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair. Mut. Res. 462, 129-135.

Lindahl, T., and Nyberg, B. (1974). Heat-induced deamination of cytosin residues in deoxyribonucleicacid. Biochemistry 13, 3405-3410.

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry 11, 3610-3618.

Lutsenko, E., and Bhagwat, A. S. (1999). The role of the Escherichia coli mug protein in the removal of uracil and 3,N(4)-ethenocytosine from DNA. J. Biol. Chem. 274, 31034-31038.

Marion, M. J., and Boivin-Angele, S. (1999). Vinyl chloride-specific mutations in humans and animals. IARC Sci. Publ. 150, 315-324.

Matsumoto, Y., and Kim, K. (1995). Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269, 699-702.

Matter, B., Wang, G., Jones, R., and Tretyakova, N. (2004). Formation of Diastereomeric Benzo[a] pyrene Diol Epoxide-Guanine Adducts in p53 Gene-Derived DNA Sequences. Chem. Res. Toxicol. 17, 731-741.

McCullough, A. K., Dodson, M. L., and Lloyd, R. S. (1999). Initiation of Base Excision Repair: Glycosylase Mechanisms And Structures. Annu. Rev. Biochem. 68, 255-285.

Missero, C., Pirro, M. T., Simeone, S., Pischetola, M., and Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. J. Biol. Chem. 276, 33569-33575.

Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell 80, 869-878.

Mol, C. D., Hosfield, D. J., and Tainer, J. A. (2000). Abasic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means. Mut. Res. 460, 211-229.

Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. Nature 403, 451-456.

Mol, C. D., Kuo, C. F., Thayer, M. M., Cunningham, R. P., and Tainer, J. A. (1995). Structure and function of the multifunctional DNA-repair enzyme exonuclease III. Nature 374, 381-386.

Nair, J., Barbin, A., Guichard, Y., and Bartsch, H. (1995). 1,N6-Ethenodeoxyadenosine and 3,N4-ethenodeoxycytidine in liver DNA from humans and untreated rodents detected by immunoaffinity/32P-postlabelling. Carcinogenesis 16, 613-617.

Nair, J., Barbin, A., Velic, I., and Bartsch, H. (1999). Etheno DNA-base adducts from endogenous reactive species. Mut. Res. 424, 59-69.

Nakamura, J., Walker, V. E., Upton, P. B., Chiang, S. Y., Kow, Y. W., and Swenberg, J. A. (1998). Highly sensitive apurinic/apyrimidinic site assay can detect spontaneous and chemically induced depurination under physiological conditions. Cancer Res. 58, 222-225.

Nash, R. A., Caldecott, K. W., Barnes, D. E., and Lindahl, T. (1997). XRCC1 protein interacts with one of two distinct forms of DNA ligase III. Biochemistry 36, 5207-5211.

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., and Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. J. Biol. Chem. 271, 12767-12774.

Neddermann, P., and Jiricny, J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. Proc. Natl. Acad. Sci. U. S. A. 91, 1642-1646.

Neuberger, M. S., Harris, R. S., Noia, J. D., and Petersen-Mahrt, S. K. (2003). Review; Immunity through DNA deamination. Trends Biochem. Sci. 28, 305-312.

Ng, H.-H., and Bird, A. (1999). DNA methylation and chromatin modification. Curr. Opin. Genet. Dev. 9, 158-163.

Niederreither, K., Harbers, M., Chambon, P., and Dolle, P. (1998). Expression of T:G mismatch-specific thymidine-DNA glycosylase and DNA methyl transferase genes during development and tumorigenesis. Oncogene 17, 1577-1585.

Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., and Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. EMBO J. 20, 4278-4286.

Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T. A., Skorpen, F., and Krokan, H. E. (1997). Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. Nucleic Acids Res. 25, 750-755.

Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. Mol. Cell 5, 1059-1065.

O'Neill, R. J., Vorob'eva, O. V., Shahbakhti, H., Zmuda, E., Bhagwat, A. S., and Baldwin, G. S. (2003). Mismatch Uracil Glycosylase from Escherichia coli A GENERAL MISMATCH OR A SPECIFIC DNA GLYCOSYLASE? J. Biol. Chem. 278, 20526-20532.

Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, P. A., Steinsbekk, K., Bakke, O., and Krokan, H. E. (1999). Post-replicative base excision repair in replication foci. EMBO J. 18, 3834-3844.

Page, F. L., Randrianarison, V., Marot, D., Cabannes, J., Perricaudet, M., Feunteun, J., and Sarasin, A. (2000). BRCA1 and BRCA2 Are Necessary for the Transcription-Coupled Repair of the Oxidative 8-Oxoguanine Lesion in Human Cells. Cancer Res. 60, 5548-5552.

Palmer, T. (1995). Understanding ENZYMES, Fourth Edition Edition: Prentice Hall/Ellis Horwood.

Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J. 17, 5214-5226.

Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature 418, 99-104.

Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Genuardi, M., Karbowski, M., Yeung, A. T., Matsumoto, Y., and Bellacosa, A. (2000). Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. J. Cell. Physiol. 185, 473-480.

Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., Yeung, A. T., Matsumoto, Y., and Bellacosa, A. (2000). Biphasic Kinetics of the Human DNA Repair Protein MED1 (MBD4), a Mismatch-specific DNA N-Glycosylase. J. Biol. Chem. 275, 32422-32429.

Pogozelski, W. K., and Tullius, T. D. (1998). Oxidative Strand Scission of Nucleic Acids: Routes Initiated by Hydrogen Abstraction from the Sugar Moiety. Chem. Rev. 98, 1089-1108.

Prasad, R., Singhal, R. K., Srivastava, D. K., Molina, J. T., Tomkinson, A. E., and Wilson, S. H. (1996). Specific interaction of DNA polymerase beta and DNA ligase I in a multiprotein base excision repair complex from bovine testis. J. Biol. Chem. 271, 16000-16007.

Privezentzev, C. V., Saparbaev, M., and Laval, J. (2001). The HAP1 protein stimulates the turnover of human mismatch-specific thymine-DNA-glycosylase to process 3,N(4)-ethenocytosine residues. Mut. Res. 480-481, 277-284.

Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., and Neuberger, M. S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr. Biol. 12, 1748-1755.

Razin, A., Cedar, H., and Riggs, A. D. (1984). DNA Methylation; Biochemistry and Biological Significance: Springer-Verlag New York Inc.).

Razin, A., and Riggs, A. D. (1980). DNA methylation and gene function. Science 210, 604-610.

Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Ruggeri, B. A., Klein-Szanto, A. J. P., Testa, J. R., Neri, G., and Bellacosa, A. (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nature Genet. 23, 266-268.

Rothwell, D. G., Hang, B., Gorman, M. A., Freemont, P. S., Singer, B., and Hickson, I. D. (2000). Substitution of Asp-210 in HAP1 (APE/Ref-1) eliminates endonuclease activity but stabilises substrate binding. Nucleic Acids Res. 28, 2207-2213.

Sagi, J., Perry, A., Hang, B., and Singer, B. (2000). Differential destabilization of the DNA oligonucleotide double helix by a T.G mismatch, 3,N(4)-ethenocytosine, 3,N(4)-ethanocytosine, or an 8-(hydroxymethyl)-3,N(4)-ethenocytosine adduct incorporated into the same sequence contexts. Chem. Res. Toxicol. 13, 839-845.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). 3 Molecular Cloning A Laboratory Manual, Second Edition Edition: Cold Spring Harbor Laboratory Press.

Saparbaev, M., and Laval, J. (1998). 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. Proc. Natl. Acad. Sci. U. S. A. 95, 8508-8513.

Sard, L., Tornielli, S., Gallinari, P., Minoletti, F., Jiricny, J., Lettieri, T., Pierotti, M. A., Sozzi, G., and Radice, P. (1997). Chromosomal localizations and molecular analysis of TDG gene-related sequences. Genomics 44, 222-226.

Scharer, O. D. (2003). Chemistry and Biology of DNA Repair. Angew. Chem. Int. Ed. 42, 2946-2974.

Scharer, O. D., Kawate, T., Gallinari, P., Jiricny, J., and Verdine, G. L. (1997). Investigation of the mechanisms of DNA binding of the human G/T glycosylase using designed inhibitors. Proc. Natl. Acad. Sci. U. S. A. 94, 4878-4883.

Schmutte, C., Baffa, R., Veronese, L. M., Murakumo, Y., and Fishel, R. (1997). Human thymine-DNA glycosylase maps at chromosome 12q22-q24.1: a region of high loss of heterozygosity in gastric cancer. Cancer Res. 57, 3010-3015.

Seeberg, E., Eide, L., and Bjørås, M. (1995). The base excision-repair pathway. Trends Biochem. Sci. 20, 391-397.

Shibutani, S., Suzuki, N., Matsumoto, Y., and Grollman, A. P. (1996). Miscoding properties of 3,N4-etheno-2'-deoxycytidine in reactions catalyzed by mammalian DNA polymerases. Biochemistry 35, 14992-14998.

Shimizu, Y., Iwai, S., Hanaoka, F., and Sugasawa, K. (2003). Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. Embo J. 22, 164-173.

Sibghat-Ullah, and Day, R. S., III (1995). Site specificity of incisions at G:T and O6-methylguanine:T base mismatches in DNA by human cell-free extracts. Biochemistry 34, 6869-6875.

Sibghat-Ullah, Gallinari, P., Xu, Y.-Z., Goodman, M. F., Bloom, L. B., Jiricny, J., and Day, R. S., III (1996). Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase. Biochemistry 35, 12926-12932.

Singer, B., and Hang, B. (1999). Mammalian enzymatic repair of etheno and para-benzoquinone exocyclic adducts derived from the carcinogens vinyl chloride and benzene. IARC Sci. Publ. 150, 233-147.

Strauss, P.R., Beard, W.A., Patterson, T.A., and Wilson, S.H. (1997) Substrate Binding by Human Apurinic/Apyrimidinic Endonuclease Indicates a Briggs-Haldane Mechanism, *J. Biol. Chem.*, **272** (2), 1302-1307

Sussel, L., Marin, O., Kimura, S., and Rubenstein, J. L. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. Development 126, 3359-3370.

Tini, M., Benecke, A., Um, S. J., Torchia, J., Evans, R. M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol. Cell 9, 265-277.

Umar, A., Buermeyer, A. B., Simon, J. A., Thomas, D. C., Clark, A. B., Liskay, R. M., and Kunkel, T. A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. Cell 87, 65-73.

Vidal, A. E., Boiteux, S., Hickson, I. D., and Radicella, J. P. (2001). XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. Embo J. 20, 6530-6539.

Vidal, A. E., Hickson, I. D., Boiteux, S., and Radicella, J. P. (2001). Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step. Nucleic Acids Res. 29, 1285-1292.

Vogelstein, B. (1990). A deadly inheritance. Nature 348, 681-682.

Walker, L. J., Robson, C. N., Black, E., Gillespie, D., and Hickson, I. D. (1993). Identification of residues in the human DNA repair enzyme HAP1 (Ref-1) that are essential for redox regulation of Jun DNA binding. Mol. Cell Biol. 13, 5370-5376.

Waters, T. R., Gallinari, P., Jiricny, J., and Swann, P. F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. J. Biol. Chem. 274, 67-74.

Waters, T. R., and Swann, P. F. (1998). Kinetics of the action of thymine DNA glycosylase. J. Biol. Chem. 273, 20007-20014.

Watson, J. D., and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737-738.

Werner, R. M., and Strivers, J. T. (2000). Kinetic Isotope Effect Studies of the Reaction Catalyzed by Uracil DNA Glycosylase: Evidence for an Oxocarbenium Ion-Uracil Anion Intermediate. Biochemistry 39, 140054-14064.

Wibley, J. E. A., Waters, T. R., Haushalter, K., Verdine, G. L., and Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. Mol. Cell 11, 1647-1659.

Wiebauer, K., and Jiricny, J. (1989). In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. Nature 339, 234-236.

Willson III, D. M., Takeshita, M., Grollman, A. P., and Demple, B. (1995). Incision Activity of Human Apurinic Endonuclease (Ape) at Abasic Site Analogs in DNA. J. Biol. Chem. 270, 16002-16007.

Wilson, S. H., and Kunkel, T. A. (2000). Passing the baton in base excision repair. Nat. Struct. Biol. 7, 176-178.

Wood, R. D. (1997). Nucleotide excision repair in mammalian cells. J. Biol. Chem. 272, 23465-23468.

Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). Human DNA Repair Genes. Science 291, 1284-1289.

Xanthoudakis, S., Miao, G., Wang, F., Pan, Y. C., and Curran, T. (1992). Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. Embo J. 11, 3323-3335.

Xu, Y.-Z. and Swann, P.F. (1992). Chromatographic separation of oligodeoxynucleotides with identical length: application to purification of oligomers containing a modified base. Anal. Biochem. 204, 185-189

Yang, H., Clendenin, W. M., Wong, D., Demple, B., Slupska, M. M., Chiang, J. H., and Miller, J. H. (2001). Enhanced activity of adenine-DNA glycosylase (Myh) by apurinic/apyrimidinic endonuclease (Ape1) in mammalian base excision repair of an A/GO mismatch. Nucleic Acids Res. 29, 743-752.

Yoon, J.-H., Iwai, S., O'Connor, T. R., and Pfeifer, G. P. (2003). Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. Nucleic Acids Res. 31, 5399-5404.

Zhang, W., Johnson, F., Grollman, A. P., and Shibutani, S. (1995). Miscoding by the exocylic and and related DNA adducts 3,N4-etheno-2'-deoxycytidine, 3,N4-ethano-2'-deoxycytidine, and 3-(2-hydroxy-ethyl)-2'-deoxyuridine. Chem. Res. Toxicol. 8, 157-163.

Zhang, W., Rieger, R., Iden, C., and Johnson, F. (1995). Synthesis of 3,N4-etheno, 3,N4-ethano, and 3-(2-hydroxyethyl) derivatives of 2'-deoxycytidine and their incorporation into oligomeric DNA. *Chem. Res. Toxicol.* 8, 148-156.

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., and Thiry, S. (2000). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res. 28, 4157-4165.